1268768

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

January 04, 2005

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *60/523,609*
FILING DATE: *November 19, 2003*
RELATED PCT APPLICATION NUMBER: *PCT/US04/39066*

UNITED STATES PATENT AND TRADEMARK OFFICE
DEPARTMENT OF COMMERCE

Certified By

Jon W Dudas

Under Secretary
of Commerce for Intellectual Property
and Acting Director of the
Unites States Patent and Trademark Office

05100
PATENT TRADEMARK OFFICE

PTO/SB/16(6-95)
Approved for use through 04/11/98. OMB0651-0037
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

PROVISIONAL APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION under 37 C.F.R. § 1.53(b)(2)

03917 U.S.PTO 60/523609 111903

Express Mail label number EV 299 896 440 US Date of Deposit: November 19, 2003
I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. § 1.10 on the date indicated above and is addressed to MS Provisional Patent Application, the Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

Richard T. Ito
Name of person signing

Signature

Docket Number	GC819P	Type a plus sign (+) inside this box →	+

INVENTOR(s)/APPLICANT(s)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
Jones	Brian	E.	2263 VA, Leidschendam
Poulose	Ayrookaran	J.	Belmont, CA 94022

TITLE OF THE INVENTION (280 characters max)

SERINE PROTEASES, NUCLEIC ACIDS ENCODING SERINE ENZYMES AND VECTORS AND HOST CELLS INCORPORATING SAME

CORRESPONDENCE ADDRESS

GENENCOR INTERNATIONAL, INC.
925 Page Mill Road
Palo Alto, California 94304-1013
Telephone: (650) 846-4020
Facsimile: (650) 845-6504

ENCLOSED APPLICATION PARTS (*check all that apply*)

☒ Specification ☒ Drawing(s)—Figures 1A – 32	*Number of Pages:* 82 *Number of Sheets:* 33	☐ Other (specify)_____	

METHOD OF PAYMENT (check one)

☐ A check or money order is enclosed to cover the Provisional filing fees. ☒ The Commissioner is hereby authorized to charge filing fees and credit Deposit Account Number: 07-1048.	PROVISIONAL FILING FEE AMOUNT ($)	$160.00

The PTO did not receive the following listed item(s) Only 21 pages 7 Drawings is received

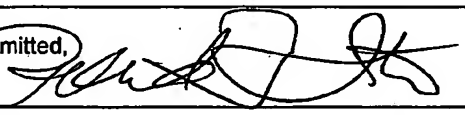
BEST AVAILABLE COPY

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☒ No.
☐ Yes, the same of the U.S. Government agency and the Government contract number are:

_____

Respectfully submitted,

SIGNATURE 

Date: November 19, 2003

TYPED or PRINTED NAME Richard T. Ito

REGISTRATION NO. 32,242
(if appropriate)

CERTIFICATE OF MAILING BY "EXPRESS MAIL"
I hereby certify that this correspondence is being deposited with the US Postal Services "Express Mail Post Office to Addressee" service under 37 CFR 1.10, Express Mail Label No. EV 299 896 440 US and addressed to Box Provisional Patent Application, Commissioner for Patents, Washington, D.C. 20231 on the date shown below:

Date: November 19, 2003 By: Richard Ito

IPATENT
Docket No. GC 819-P

SERINE PROTEASES, NUCLEIC ACIDS ENCODING SERINE ENZYMES AND VECTORS AND HOST CELLS INCORPORATING SAME

Inventors:
Brian E. Jones, Ayrookaran J. Poulouse

FIELD OF THE INVENTION

The present invention is directed to a novel serine protease, novel genetic material encoding that enzyme and proteolytic proteins developed therefrom. In particular, the present invention provides a protease derived from a *Cellulomonas* spp, a DNA encoding such protease, vectors comprising such DNA, host cells transformed with such DNA and an enzyme produced by such host cells. The present invention also provides a cleaning composition (e.g., a detergent composition), animal feed composition, textile and leather processing compositions comprising the protease derived from a *Cellulomonas* spp.

BACKGROUND OF THE INVENTION

Serine proteases are a subgroup of carbonyl hydrolases. They comprise a diverse class of enzymes having a wide range of specificities and biological functions. Stroud, R. Sci. Amer., 131:74-88. Despite their functional diversity, the catalytic machinery of serine proteases has been approached by at least two genetically distinct families of enzymes: 1) the subtilisins and 2) the mammalian chymotrypsin-related and homologous bacterial serine proteases (e.g., trypsin and *S. griseus* trypsin). These two families of serine proteases show remarkably similar mechanisms of catalysis. Kraut, J. (1977), Annu. Rev. Biochem., 46:331-358. Furthermore, although the primary structure is unrelated, the tertiary structure of these two enzyme families brings together a conserved catalytic triad of amino acids consisting of serine, histidine and aspartate. The subtilisins and chymotrypsin related serine proteases both have a catalytic triad comprising aspartate, histidine and serine. In the subtilisin related proteases the relative order of these amino acids, reading from the amino to carboxy terminus, is aspartate-histidine-serine. In the chymotrypsin related proteases, the relative order, however, is histidine-aspartate-serine. Much has been said and explored

with regard to the subtilisins. In particular, subtilisins are useful in cleaning and feed applications. However, use in these applications often subject the subtilisins to adverse environmental conditions, e.g., exposure to oxidative agents, chelating agents, adverse thermal conditions, and/or adverse pH conditions, which can adversely affect the activity of the protease in use

Streptomyces grows as an extensively branching primary or substrate mycelium and an abundant aerial mycelium that at maturity bear characteristic spores. Streptogrisins are serine proteases (approx. MW about 18,000 to about 19,000) which are secreted in large amounts from a wide variety of *Streptomyces* species. The amino acid sequences of streptomyces proteases has been determined from at least 9 different species of *Streptomyces (Streptomyces griseus* Streptogrisin C (accession no. P52320); alkaline proteinase (EC 3.4.21.-) from *Streptomyces* sp. (accession no. PC2053); alkaline serine proteinase I from *Streptomyces* sp. (accession no. S34672), serine protease from *Streptomyces lividans* (accession no. CAD4208); putative serine protease from *Streptomyces coelicolor* A3(2) (accession no. NP_625129); putative serine protease from *Streptomyces avermitilis* MA-4680 (accession no. NP_822175); serine protease from *Streptomyces lividans* (accession no. CAD42809); putative serine protease precursor from *Streptomyces coelicolor* A3(2) (accession no. NP_628830)). A purified native alkaline protease having an apparent molecular weight of 19,000 daltons and isolated from *Streptomyces griseus* var. *alkaliphilus* protease and cleaning compositions comprised thereof are described in US Patent No. 5,646,028.

Gram-positive alkaliphilic bacteria that had been isolated from in and around alkaline soda lakes were described in US Patent No. 5,401,657. These alkali-philes were analyzed according to the principles of numerical taxonomy with respect to each other and also a collection of known bacteria.and taxonomically characterized. Six natural clusters or phenons of alkaliphilic bacteria were generated. Amongst the strains isolated was a strain identified as 69B4.

Cellulomonas spp. are Gram positive bacteria. *Cellulomonas* grows as slender, often irregular rods that may occasionally show branching, but no mycelium is formed. There is no aerial growth and no spores are formed. *Cellulomonas* and *Streptomyces* are only distantly related at a genetic level. The large genetic (genomic) distinction between *Cellulomonas* and *Streptomyces* is reflected in a great difference in phenotypic properties. While serine protease in Streptomyces have been previously examined, the inventors are unaware of any previous description of serine protease (approx. MW 18,000 to 20,000)

which are secreted by *Cellulomonas* spp. The inventors are unaware of any known instances of *Cellulomonas* proteases being used in the cleaning and/or feed industry.

It is, thus, generally desirable to discover and develop novel enzymes having good stability and proteolytic activity for use in connection with cleaning compositions, animal feed, and to apply advancements in fermentation technology to the production of such enzymes in order to make them commercially viable. It is also desirable to ascertain nucleotide sequences which can be used to produce more efficient genetically engineered organisms capable of expressing such hydrolytic proteases in quantities suitable for industrial production. It is still further desirable to develop a *Cellulomonas* serine protease expression system via genetic engineering which will enable the purification and utilization of working quantities of relatively pure enzyme.

SUMMARY OF THE INVENTION

According to the present invention, an isolated polypeptide having proteolytic activity, e.g., a protease, is provided comprising at least 70% amino acid sequence identity to SEQ ID NO. 8. In another aspect an isolated polypeptide having proteolytic activity, e.g., a protease, is provided comprising at least 95% amino acid sequence identity to SEQ ID NO.:8. In another aspect the isolated polypeptide having proteolytic activity, e.g., a protease, is provided comprising an amino acid sequence SEQ ID NO.:8. In one aspect, the polypeptide, e.g., protease, is derived from *Cellulomonas spp.* In another aspect, the *Cellulomonas spp.* is selected from *Cellulomonas fimi*, *Cellulomonas biazotea*, *Cellulomonas cellasea*, *Cellulomonas homis*, *Cellulomonas flavigena*, *Cellulomonas persica*, *Cellulomonas iranensis*, *Cellulomonas gelida*, *Cellulomonas humilata*, *Cellulomonas turbata*, *Cellulomonas uda*, and *Cellulomonas* strain 69B4 (DSM 16035). In one aspect the protease has an apparent molecular weight of about 17kD to 21kD as determined by a matrix assisted laser desorption/ionizaton – time of flight ("MALDI-TOF") spectrophotometer.

According to the present invention, an isolated polynucleotide is provided encoding a protease comprising an amino acid sequence comprising at least 70% amino acid sequence identity to SEQ ID NO.8. In one aspect of the invention, an isolated polynucleotide is provided encoding a protease comprising an amino acid sequence having at least 95% amino acid sequence sequence identity to SEQ.ID.NO.:8. In another aspect of the invention, an isolated polynucleotide is provided encoding a protease comprising an amino acid sequence of SEQ.ID.NO.:8. According to the present invention, a vector is provided comprising a polynucleotide described above, e.g., encoding a protease having at least 70%

sequence identity to SEQ ID. NO.8, having at least 95% seguence identity to SEQ ID. NO.8, and/or having an amino acid sequence of SEQ ID. NO. 8. According to the present invention, a host cell is provided that is transformed with such vector. According to the present invention an enzyme composition is provided comprising the novel protease expressed by the host cell. According to the present invention, a cleaning composition is provided comprising the novel protease.

According to the present invention, an isolated polynucleotide encoding a protease is provided comprising at least 70% nucleotide sequence identity to SEQ ID NO.4. In one aspect of the invention, an isolated polynucleotide encoding a protease is provided comprising at least 95% sequence identity to SEQ.ID.NO.:4. In another aspect of the invention, an isolated polynucleotide encoding a protease is provided comprising SEQ.ID.NO.:4. According to the present invention, a vector is provided comprising a polynucleotide described above, e.g., having at least 70% nucleotide sequence identity to SEQ ID. NO.4, having at least 95% nucleotide sequence identity to SEQ ID. NO.4, and/or comprising SEQ ID. NO.:4. According to the present invention, a host cell is provided that is transformed with such vector. According to the present invention an enzyme composition is provided comprising the novel protease expressed by the host cell. According to the present invention, a cleaning composition is provided comprising the novel protease.

According to the present invention, an isolated polynucleotide is provided comprising a nucleotide sequence (i) having at least 70% identity to SEQ ID NOS.:3 or 4, or (ii) being capable of hybridizing to a probe derived from the nucleotide sequence disclosed in SEQ ID NOS: 3 or 4, under conditions of medium to high stringency, or (iii) being complementary to the nucleotide sequence disclosed in SEQ ID NOS:: 3 or 4. According to the present invention, a vector is provided comprising such polynucleotide. According to the present invention, a host cell is provided that is transformed with such vector.

According to the present a method of producing an enzyme having protease activity is provided comprising: the steps of transforming a host cell with an expression vector comprising a polynucleotide comprising at least 70% sequence identity to SEQ ID NO.4, cultivating said transformed host cell under conditions suitable for said host cell to produce said protease; and recovering said protease. According to the present invention, the host cell is a *Streptomyces* spp, a *Bacillus* spp, a *Trichoderma* spp. and/or a *Aspergillus* spp. In another aspect, the *Streptomyces* spp. is *Streptomyces lividans*. In another aspect the host cell is *Trichoderma resei*. In another aspect, the *Aspergilllus* spp. is *Aspergillus niger*.

According to the present invention a probe is provided comprising a polynucleotide

substantially identical to a fragment of SEQ ID NOS. 1, 2, 3 or 4, wherein said probe is used to detect a nucleic acid sequence coding for an enzyme having proteolytic activity, and wherein said nucleic acid sequence is obtained from a bacterial source. In one aspect the bacterial source is a *Cellulomonas* spp. In one aspect of the invention, the bacterial source is *Cellulomonas* strain 69B4.

According to the present invention, a cleaning composition is provided comprising a cleaning effective amount of a protease comprising an amino acid sequence having at least 70 % sequence identity to SEQ ID NO:8, at least 95 % sequence identity to SEQ ID NO:8, and/or having an amino acid sequence of SEQ ID NO:8; and a suitable cleaning adjunct. In another aspect, the protease is derived from a *Cellulomonas* sp. In another aspect, the *Cellulomonas* spp. is selected from *Cellulomonas fimi*, *Cellulomonas biazotea*, *Cellulomonas cellasea*, *Cellulomonas hominis*, *Cellulomonas flavigena*, *Cellulomonas persica*, *Cellulomonas iranensis*, *Cellulomonas gelida*, *Cellulomonas humilata*, *Cellulomonas turbata*, *Cellulomonas uda*, and *Cellulomonas* strain 69B4 (DSM 16035). In another aspect of the present invention, the *Cellulomonas* spp is *Cellulomonas*. strain 69B4. In another aspect, the cleaning composition further comprises one or more additional enzymes or enzyme derivatives selected from the group consisting of protease, amylase, lipase, mannanase and cellulase.

In another aspect of the present invention, an isolated naturally occurring protease is provided comprising an amino acid sequence having at least 70% sequence identity to SEQ ID NO: 8, at least 95% sequence identity to SEQ ID NO: 8, and/or having the sequence identity of SEQ ID NO: 8, the protease being isolated from a *Cellulomonas* spp.. In another aspect of the invention, the protease is isolated from *Cellulomonas* strain 69B4 (DSM 16035).

In another aspect of the present invention, a polynucleotide is provided encoding a signal peptide (i) having at least 70% sequence identity to SEQ ID NO.:9, or (ii) being capable of hybridizing to a probe derived from the polypeptide sequence encoding SEQ ID NO: 9, under conditions of medium to high stringency, or (iii) being complementary to the polypeptide sequence disclosed in SEQ ID NO:9. In another aspect of the present invention a vector is provided comprising the polynucleotide described above. In another aspect of the present invention, a host cell is provided that is transformed with the vector.

Another aspect of the present invention provides a method of producing a protease, comprising:

(a) transforming a host cell with an expression vector comprising a polynucleotide having at least 70% sequence identity to SEQ.ID.NO.4, at least 95% sequence identity to SEQ.ID.NO.4, and/or having a polynucleotide sequence of SEQ.ID.NO.4;

(b) cultivating the transformed host cell under conditions suitable for the host cell to produce the protease; and

(c) recovering the protease.

According to one aspect, the host cell is a *Bacillus* species, for example *B. subtilis.* In another aspect, the host cell is a *Streptomyces spp.*, for example *Streptomyces lividans.* In another aspect the host cell is a *Trichoderma spp.*, for example *Trichoderma resei.* In another aspect the host cell is a *Aspergillus spp.*, for example *Aspergillus niger.*

As will be appreciated, an advantage of the present invention is that a polynucleotide has been isolated which provides the capability of isolating further polynucleotides which encode proteins having serine protease activity, said backbone substantially identical to that of the *Cellulomonas* protease of the invention.

Another advantage of the present invention is that, by virtue of providing a polynucleotide encoding a protein having serine protease activity, it is possible to produce through recombinant means a host cell that is capable of producing the protein having serine protease activity in relatively large quantities.

Yet another advantage of the present invention is that commercial application of proteins having serine protease activity is made practical. For example, the present invention provides animal feed incorporating the serine protease described herein.

Still a further advantage of the present invention is that it provides an enzyme composition having proteolytic activity, with such activity being suitable for use in applications that benefit from proteolytic activity, for example, but not limited to, animal feed, textile processing, leather finishing, grain processing, meat processing, cleaning, preparation of protein hydrolysates, personal care products, including oral care and/or skin care.

Still a further advantage of the present invention is that it provides an enzyme having comparative wash performance with presently utilized *subtilisin* proteases, which make it very suitable to for use in detergents or cleaning compositions.

Other objects and advantages of the present invention will become apparent from the following detailed specification.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGS. 1A-B show the DNA sequence of the *asp* gene (SEQ ID NO.1) that encodes for, e.g., the signal peptide (SEQ ID NO:9) and the precursor serine protease (SEQ ID NO.:8) derived from *Cellulomonas strain 69B4* (DSM 16035). The initiating polynucleotide encoding the signal peptide of the *Cellulomonas* strain 69B4 protease is in bold (ATG).

FIGS. 2A-B show the DNA sequence (SEQ ID NO.:2) encoding the signal peptide (SEQ ID NO:9) that is operatively linked to the serine protease (SEQ ID NO.:8) and the precursor protease (SEQ ID NO.:7) derived from *Cellulomonas strain 69B4* (DSM 16035). The initiating polynucleotide encoding the signal peptide of the *Cellulomonas* strain 69B4 protease is in bold (ATG). The asterisk indicates the termination codon, beginning with residue 1486. Residues 85, 595, and 1162, relate to the initial residues of the N terminal prosequence, mature sequence and Carbon terminal prosequence, respectively, are bolded and underlined.

FIGS. 3A-B show the DNA sequence (SEQ ID NO.:3) encoding the precursor protease derived from *Cellulomonas* strain 69B4 (DSM 16035).

FIG. 4 show the DNA sequence (SEQ ID NO.:4) encoding the mature protease derived from *Cellulomonas* strain 69B4 (DSM 16035).

FIG. 5 show the DNA sequence (SEQ ID NO.:5) encoding the signal peptide derived from *Cellulomonas* strain 69B4 (DSM 16035).

FIG. 6 shows the amino acid sequence (SEQ ID NO.:6) of the signal sequence and precursor protease derived from *Cellulomonas* strain 69B4 (DSM 16035), including the signal sequence [segments 1a-c] (residues 1-28 [-198 to –171]), an N-terminal prosequence [segments 2a-r] (residues 29-198 [-170 to –1]), a mature protease [segments 3a-t] (residues 199-387 [1-189]), and a C-terminal prosequence [segments 4a-l] (residues 388-495 [190-398]) encoded by the DNA sequence of FIGS 1, 2, 3 and/or 4 (SEQ ID NO.:1, 2, 3 and/or 4). The N-terminal sequence of the mature protease amino acid sequence is in bold.

FIG. 7 shows the amino acid sequence of the precursor protease derived from *Cellulomonas strain 69B4* (DSM 16035) (SEQ ID NO.:7).

FIG. 8 shows the amino acid sequence of the mature protease derived from *Cellulomonas strain 69B4* (DSM 16035) (SEQ ID NO.:8). Catalytic triad H32, D56 and S132 are bolded and underlined.

FIG. 9 shows the amino acid sequence of the signal peptide of the protease derived from *Cellulomonas strain 69B4* (DSM 16035) (SEQ ID NO.:9).

FIG 10 shows a phylogenetic tree of *Cellulomonas* spp. based on 16s RNA gene

sequence identity. Bar = 2% sequence divergence.

FIG. 11 shows the degenerate primer (SEQ ID No. 10) used to identify a 177 bp fragment of the protease of *Cellulomonas* strain 69B4.

FIG. 12 shows the reverse primer (SEQ ID NO.:11) used to identity a 177 bp fragment of the protease derived from *Cellulomonas* strain 69B4.

FIG. 13 shows the DNA and amino acid sequence of the 177 bp fragment (SEQ ID NO. 12) encoding part of the protease gene derived from *Cellulomonas* strain 69B4. Sequences of the degenerate primers of FIGS. 10 and 11 are underlined and in bold.

FIGS. 14A and B shows an alignment of the amino acid sequences of *Streptomyces* serine proteases and *Thermobifida fusca* serine proteases. The accession number identifies the respective sequence. *Streptomyces griseus* Streptogrisin C (accession no. P52320); *Thermobifida fusca* (accession no. AAC23545); alkaline proteinase (EC 3.4.21.-) from *Streptomyces* sp. (accession no. PC2053); alkaline serine proteinase I from *Streptomyces* sp. (accession no. S34672), serine protease from *Streptomyces lividans* (accession no. CAD42808); putative serine protease from *Streptomyces coelicolor* A3(2) (accession no. NP 625129); putative serine protease from *Streptomyces avermitilis* MA-4680 (accession no. NP 822175); serine protease from *Streptomyces lividans* (accession no. CAD42809); putative serine protease precursor from *Streptomyces coelicolor* A3(2) (accession no. NP 628830). Two conserved boxes are underlined and in bold.

FIG. 15 shows an alignment of the amino acid sequences of *Cellulomonas* 69B4 mature protease ("69B4 mature") with mature proteases amino acid sequences of Streptogrisin C ("Sq – streptogrisinC_mature"), Streptogrisin B ("Sq – streptogrisinBmature"), Streptogrisin A ("Sq – streptogrisinAmature"), Streptogrisin D ("Sq – streptogrisinDmature") and consensus residues.

FIG. 16 shows an amino acid sequence alignment of the *Cellulomonas* 69B4 signal sequence and precursor protease with the signal sequence and precursor protease Aqualysin I of *Thermus aquaticus* (COOH-terminal pro-sequence of Aqualysin I is underlined and in bold).

FIG. 17 shows a MALDI TOF spectrum of a protease derived from Cellulomonas strain 69B4

FIG. 18 shows the construction of pSEGCT69B4.

FIG. 19 compares the cleaning activity [absorbance at 405 nm] dose [ppm] response curves of certain serine proteases (69B4 [-x-]; PURAFECT [-◊-]; RELASE [-Δ-]; and OPTIMASE [-□-] in liquid TIDE brand detergent (Procter & Gamble, Cincinnati, OH, USA)

under North American wash conditions.

FIG. 20 compares the cleaning activity [absorbance at 405 nm] dose [ppm] response curves of certain serine proteases (69B4 [-x-]; PURAFECT [-◇-]; RELASE [-Δ-]; and OPTIMASE [-□-] in TIDE OPAL brand powder detergent (Procter & Gamble, Cincinnati, OH, USA) 0.66 g/l (North American concentration/detergent formulation) under Japanese wash conditions.

FIG. 21 compares the cleaning activity [absorbance at 405 nm] dose [ppm] response curves of certain serine proteases (69B4 [-x-]; PURAFECT [-◇-]; RELASE [-Δ-]; and OPTIMASE [-□-] in ARIEL REGULAR brand detergent powder (Procter & Gamble, Cincinnati, OH, USA) under European wash conditions.

FIG. 22 compares the cleaning activity [absorbance at 405 nm] dose [ppm] response curves of certain serine protease (69B4 [-x-]; PURAFECT [-◇-]; RELASE [-Δ-]; and OPTIMASE [-□-] in PURE CLEAN brand detergent powder (Procter & Gamble, Cincinnati, OH, USA) under Japanese conditions.

FIG. 23 compares the cleaning activity [absorbance at 405 nm] dose [ppm] response curves of certain serine proteases (69B4 [-x-]; PURAFECT [-◇-]; RELASE [-Δ-]; and OPTIMASE [-□-] in Tidal Opal brand detergent powder (Procter & Gamble, Cincinnati, OH, USA) 1.00 g/l under North American conditions.

FIG. 24 shows comparative oxidative inactivation of various serine proteases (100 ppm) as a measure of per cent enzyme activity over time (minutes) (69B4 [-x-]; BPN' variant 1 [-◇-]; PURAFECT [-Δ-]; and GG36-variant 1 [-□-]) with 0.1 M H_2O_2 at pH 9.45, 25° C.

FIG. 25 shows comparative chelator inactivation of various serine proteases (100 ppm) as a measure of per cent enzyme activity over time (minutes) (69B4 [-x-]; BPN'-variant 1 [-◇-]; PURAFECT [-Δ-]; and GG36-variant 1 [-□-] with 10mM EDTA at pH 8.20, 45° C.

FIG. 26 shows comparative thermal inactivation of various serine proteases (100 ppm) as a measure of per cent enzyme activity over time (minutes) (69B4 [-x-]; BPN'-variant [-◇-]; PURAFECT [-Δ-]; and GG36-variant 1 [-□-] with 50 mM Tris at pH 8.0, 45° C.

FIG. 27 shows comparative thermal inactivation of certain serine proteases (69B4 [-x-]; BPN'-variant [-◇-]; PURAFECT [-Δ-]; and GG36-variant-1 [-□-] at pH 8.60, over a temperature gradient of 57° C. - 62° C.

FIG. 28 shows enzyme activity (hydrolysis of di-methyl casein measured by absorbance at 405 nm) of certain serine proteases (2.5 ppm) (69B4 [-□-]; BPN'-variant [-◇- PURAFECT [-Δ-]; and GG36-variant 1[-○ -] at pH 's ranging from 5 to 12 at 37° C.

FIG. 29 is a bar graph which shows enzyme stability as indicated by % remaining activity (hydrolysis of di-methyl casein measured by absorbance at 405 nm) of certain serine proteases (2.5 ppm) (69B4, BPN'- variant; PURAFECT and GG36-variant 1 at pH 's ranging from 3 (☐), 4 (▨), 5 (▤) to 6 (▧) at 25 °, 35° and 45° C., respectively.

FIG 30 shows enzyme stability as indicated by % remaining activity of a BPN'-variant at pH ranges from 3 (-◇-), 4 (--□--), 5 (--Δ--) to 6 (--X--) at 25 °, 35° and 45° C., respectively

FIG 31 shows enzyme stability as indicated by % remaining activity of PURAFECT TM protease at pH ranges from 3 (-◇-), 4 (--□--), 5 (--Δ--) to 6 (--X--) at 25 °, 35° and 45° C., respectively

FIG 32 shows enzyme stability as indicated by % remaining activity of 69B4 protease at pH ranges from 3 (-◇-), 4 (--□--), 5 (--Δ--) to 6 (--X--) at 25 °, 35° and 45° C., respectively

DETAILED DESCRIPTION OF THE INVENTION

I. Definitions

Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Singleton, *et al.*, DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY, 2D ED., John Wiley and Sons, New York (1994), and Hale & Marham, THE HARPER COLLINS DICTIONARY OF BIOLOGY, Harper Perennial, NY (1991) provide one of skill with a general dictionary of many of the terms used in this invention. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. Numeric ranges are inclusive of the numbers defining the range. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively. The headings provided herein are not limitations of the various aspects or embodiments of the invention which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification as a whole.

The terms "polynucleotide" and "nucleic acid", used interchangeably herein, refer to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. These terms include, but are not limited to, a single-, double- or

triple-stranded DNA, genomic DNA, cDNA, RNA, DNA-RNA hybrid, or a polymer comprising purine and pyrimidine bases, or other natural, chemically, biochemically modified, non-natural or derivatized nucleotide bases. The following are non-limiting examples of polynucleotides: a gene or gene fragment, a chromosomal fragment, EST's, exons, introns, mRNA, tRNA, rRNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs, uracyl, other sugars and linking groups such as fluororibose and thioate, and nucleotide branches. The sequence of nucleotides may be interrupted by non-nucleotide components.

The term or phrase "protease" or "proteolytic activity" refers to a protein or peptide exhibiting the ability to hydrolyze peptides or substrates having peptide linkages. Many well known procedures exist for measuring proteolytic activity (K. M. Kalisz, "Microbial Proteinases," Advances in Biochemical Engineering/Biotechnology, A. Fiechter ed., 1988). For example, proteolytic activity may be ascertained by comparative assays which analyze the respective protease's ability to hydrolyze a commercial substrate. Exemplary substrates useful in the such analysis of protease or protelytic activity, include, but are not limited to di-methyl casein (Sigma C-9801), bovine collagen (Sigma C-9879), bovine elastin (Sigma E-1625), and bovine keratin (ICN Biomedical 902111). Colormetric assays utilizing these substrates are described in PCT Publication No. WO 99/34011 (published July 8, 1999), and US Pat. No. 6,376,450 to Ghosh, et al, which are expressly incorporated by reference herein. The pNA assay (DelMar, E. G., C. Largman, J. W. Brodrick and M. C. Geokas, ANAL. BIOCHEM., Vol. 99, pp. 316-320, (1979), incorporated herein by reference) can be used to determine the active enzyme concentration for fractions collected during gradient elution. This assay measures the rate at which *p*-nitroaniline is released as the enzyme hydrolyzes the soluble synthetic substrate, succinyl-alanine-alanine-proline-phenylalanine-*p*-nitroanilide (sAAPF-*p*NA). The rate of production of yellow color from the hydrolysis reaction is measured at 410 nm on a spectrophotometer and is proportional to the active enzyme concentration. In addition, absorbance measurements at 280 nm can be used to determine the total protein concentration. The active enzyme/total-protein ratio gives the enzyme purity.

A polynucleotide is said to "encode" an RNA or a polypeptide if, in its native state or when manipulated by methods known to those of skill in the art, it can be transcribed and/or translated to produce the RNA, the polypeptide or a fragment thereof. The anti-sense

strand of such a nucleic acid is also said to encode the sequences.

As is known in the art, a DNA can be transcribed by an RNA polymerase to produce RNA, but an RNA can be reverse transcribed by reverse transcriptase to produce a DNA. Thus a DNA can encode a RNA and vice versa.

The term "construct" refers to a nucleic acid segment that may be single- or double-stranded. The construct may be any nucleic acid molecule of genomic DNA, synthetic DNA or RNA origin, and which may be based on a complete or partial naturally-occurring nucleotide sequence encoding the protease. The construct may optionally contain other nucleic acid segments.

A DNA construct refers to a nucleic acid sequence or polynucleotide generated recombinantly or synthetically for example by in vitro PCR or other suitable techniques. DNA constructs can be used to introduce nucleic acid into a host cell. The DNA construct can be incorporated into a plasmid, chromosome or nucleic acid fragment. The term DNA construct may be used interchangeably with DNA cassette, expression cassette and other grammatical equivalents.

The term "regulatory segment" or "regulatory sequence" or "expression control sequence" refers to a polynucleotide sequence of DNA that is operatively linked with a polynucleotide sequence of DNA that encodes the amino acid sequence of a polypeptide chain to effect the expression of the encoded amino acid sequence. The regulatory sequence can inhibit, repress, or promote the expression of the operably linked polynucleotide sequnce encoding the amino acid.

The term, "operably linked" indicates that the segments are arranged so that they function in concert for their intended purposes, e.g., transcription initiates in a promoter and proceeds through the DNA sequence coding for the protease of the present invention.

The term "homologous", in the context of nucleic acid constructs, refers to a nucleic acid sequence encoding a polypeptide, enzyme, e.g., a protease, native to the host organism in question; the nucleic acid acid sequence may be introduced into the host organism in multicopy form.

The terms "heterologous" or "exogenous" refers to nucleic acids, amino acids, peptides, polypeptides or proteins, which do not naturally occur in a particular host cell. In this context, the nucleic acids, amino acids, peptides, polypeptides or proteins may be non-native or synthetic. The term "heterologous", in the context of nucleic acid constructs, refers to a DNA sequence not expressed by the host cell in nature.

As used herein the term "gene" means a polynucleotide, e.g., a DNA segment, that

is involved in producing a polypeptide and includes regions preceding and following the coding regions as well as intervening sequences (introns) between individual coding segments (exons).

As used herein when describing polypeptides, and polynucleotides, including genes, that encode them, the term for the gene is not capitalized and is italics, i.e. *ppc*. The term for the polypeptide is generally not italicized and the first letter is capitalized, i.e. Ppc.

As used herein, the term "vector" refers to a polynucleotide construct designed to introduce nucleic acids into one or more cell types. Vectors include cloning vectors, expression vectors, shuttle vectors, plasmids, cassettes and the like. The polynucleotide construct can comprise a DNA sequence encoding the protease (precursor or mature protease) that is operably linked to a suitable prosequence (secretory, etc.) capable of effecting the expression of the DNA in a suitable host.

"Host strain" or "host cell" refers to a suitable host for an expression vector comprising DNA according to the present invention.

An enzyme is "overexpressed" in a host cell if the enzyme is expressed in the cell at a higher level that the level at which it is expressed in a corresponding wild-type cell.

The term "hybridization" includes any process by which a strand of a nucleic acid joins with a complementary nucleic acid strand through base-pairing. Thus, strictly speaking, the term refers to the ability of the complement of the target sequence to bind to a test sequence, or vice-versa.

"Hybridization conditions" refers to the degree of "stringency" of the conditions under which hybridization is measured. Hybridization conditions can be based on the melting temperature (Tm) of the nucleic acid binding complex, as taught in Berger and Kimmel (1987, Guide to Molecular Cloning Techniques, Methods in Enzymology, Vol 152, Academic Press, San Diego CA) incorporated herein by reference, and confer a defined stringency. Hybridization conditions can also be based on the washing conditions employed after hybridization as known in the art.

The terms "protein" and "polypeptide" are used interchangeability herein. The 3-letter code for amino acids as defined in conformity with the IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN) is used through out this disclosure. It is also understood that a polypeptide may be coded for by more than one nucleotide sequence due to the degeneracy of the genetic code.

A prosequence is an amino acid sequence between the signal sequence and mature protease that is necessary for the secretion of the protease. Cleavage of the pro sequence

will result in a mature active protease.

The term "signal sequence" or "signal peptide" refers to any sequence of nucleotides and/or amino acids which may participate in the secretion of the mature or precursor forms of the protein. This definition of signal sequence is a functional one, meant to include all those amino acid sequences encoded by the N-terminal portion of the protein gene, which participate in the effectuation of the secretion of protein under native conditions. They are often, but not universally, bound to the N-terminal portion of a protein or to the N-terminal portion of a precursor protein. The signal sequence may be endogenous or exogenous. The signal sequence may be that normally associated with the protein, e.g., protease, or may be from a gene encoding another secreted protein. One exemplary exogenous signal sequence comprises the first seven amino acid residues of the signal sequence from *Bacillus subtilis* subtilisin fused to the remainder of the signal sequence of the subtilisin from *Bacillus lentus* (ATCC 21536).

The term "substantially the same signal activity" refers to the signal activity, as indicated by substantially the same secretion of the protease into the fermentation medium, for example a fermentation medium protease level being at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98% of the secreted protease levels in the fermentation medium as provided by the signal sequence of SEQ ID NO. 5 and/or 9.

The term "mature" form of a protein or peptide refers to the final functional form of the protein or peptide. To exemply, a mature form of the protease of the present invention at least includes the amino acid sequence identical to residue positions 1-189 of Fig. 8 (SEQ ID. NO.8).

The term "precursor" form of a protein or peptide refers to a mature form of the protein having a prosequence operably linked to the amino or carbonyl terminus of the protein. The precursor may also have a "signal" sequence operably linked, to the amino terminus of the prosequence. The precursor may also have additional polynucleotides that are involved in post-translational activity, e.g., polynucleotides cleaved therefrom to leave the mature form of a protein or peptide.

"Wild type" refers to the activity characteristic of a host cell in which endogenous proteases are not modified by alteration of the genomic sequence encoding the protease, e.g., by insertion, deletion or substitution of the amino acids constituting the structure genome of the respective protease.

"Naturally occurring enzyme" refers to an enzyme having the unmodified amino acid sequence identical to that found in nature. Naturally occurring enzymes include native

enzymes, those enzymes naturally expressed or found in the particular microorganism.

The term "*Cellulomonas* homologues" refers to naturally occurring proteases having substantially identical amino acid sequences to the mature protease derived from *Cellulomonas* strain 69B4 or polynucleotide sequences which encode for such naturally occurring proteases, and which proteases retain the functional characteristics of a serine protease encoded by such nucleic acids.

The term "69B4 protease" refers to a naturally occurring mature protease derived from *Cellulomonas* strain 69B4 (DSM 16035) having substantially identical amino acid sequences as described in Fig. 8 (SEQ ID NO.:8).

The term "derived from" refers to, not only a protease produced or producible by a strain of the organism in question, but also a protease encoded by a DNA sequence isolated from such strain and produced in a host organism containing such DNA sequence. Additionally, the term refers to a protease which is encoded by a DNA sequence of synthetic and/or cDNA origin and which has the identifying characteristics of the protease in question. To exemplify, "proteases derived from *Cellulomonas*" refers to those enzymes having proteolytic activity which are naturally-produced by *Cellulomonas*, as well as to serine proteases like those produced by *Cellulomonas* sources but which through the use of genetic engineering techniques are produced by non-*Cellulomonas* organisms transformed with a nucleic acid encoding said serine proteases. A "derivative" within the scope of this definition will retain generally the characteristic proteolytic activity observed in the wild-type, native or parent form to the extent that the derivative is useful for similar purposes as the wild-type, native or parent form. Functional derivatives of serine protease encompass naturally occurring, synthetically or recombinantly produced peptides or peptide fragments which have the general characteristics of the serine protease of the present invention.

The term "functional derivative" refers to a derivative of a nucleic acid which has the functional characteristics of a nucleic acid which encodes serine protease. Functional derivatives of a nucleic acid which encode serine protease of the present invention encompass naturally occurring, synthetically or recombinantly produced nucleic acids or fragments and encode serine protease characteristic of the present invention. Wild type nucleic acid encoding serine proteases according to the invention include naturally occurring alleles and homologues based on the degeneracy of the genetic code known in the art.

The term "identical" in the context of two nucleic acids or polypeptide sequences refers to the residues in the two sequences that are the same when aligned for maximum correspondence, as measured using one of the following sequence comparison or analysis

algorithms.

The term "optimal alignment" refers to the alignment giving the highest percent identity score.

"Percent sequence identity", "Percent amino acid sequence identity", "Percent gene sequence identityand/or "Percent nucleic acid/polynucloetide sequence identity", with respect to two amino acid, polynucleotide and/or gene sequences, refers to the percentage of residues that are identical in the two sequences when the sequences are optimally aligned. Thus, 80% amino acid sequence identity means that 80% of the amino acids in two optimally aligned polypeptide sequences are identical.

The phrase "substantially identical" in the context of two nucleic acids or polypeptides thus refers to a polynucleotide or polypeptide that comprising at least 70% sequence identity, preferably at least 75%, preferably at least 80%, preferably at least 85%, preferably at least 90%, preferably at least 95% , preferably at least 97% , preferably at least 98% and preferably at least 99% sequence identity as compared to a reference sequence using the programs or algorithms (e.g., BLAST, ALIGN, CLUSTAL) using standard parameters. One indication that two polypeptides are substantially identical is that the first polypeptide is immunologically cross-reactive with the second polypeptide. Typically, polypeptides that differ by conservative amino acid substitutions are immunologically cross-reactive. Thus, a polypeptide is substantially identical to a second polypeptide, for example, where the two peptides differ only by a conservative substitution. Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions (e.g., within a range of medium to high stringency).

The phrase "equivalent," in this context, refers to serine proteases enzymes that are encoded by a polynucleotide capable of hybridizing to the polynucleotide having the sequence as shown in any one of Figures 1A-C under conditions of medium to maximal stringency. For example, being equivalent means that an equivalent mature serine protease comprises at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98% and/or at least 99% sequence identity to the mature *Cellulomonas* serine protease having the amino acid sequence of Fig. 8 (SEQ ID NO:8).

The term "isolated" or "purified" refers to a material that is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, the material is said to be "purified" when it is present in a particular composition in a higher or

lower concentration than exists in a naturally occurring or wild type organism or in combination with components not normally present upon expression from a naturally occurring or wild type organism. For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector, and/or such polynucleotides or polypeptides could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment. A nucleic acid or protein is said to be purified, for example, if it gives rise to essentially one band in an electrophoretic gel.

The term "isolated", when applied to a DNA sequence molecule, refers to a DNA sequence that has been removed from its natural genetic milieu and is thus free of other extraneous or unwanted coding sequences, and is in a form suitable for use within genetically engineered protein production systems. Such isolated molecules are those that are separated from their natural environment and include cDNA and genomic clones. Isolated DNA molecules of the present invention are free of other genes with which they are ordinarily associated, but may include naturally occurring 5' and 3' untranslated regions such as promoters and terminators. The identification of associated regions will be evident to one of ordinary skill in the art (see for example, Dynan and Tijan, Nature 316:774-78, 1985). The term "an isolated DNA sequence" may alternatively be termed "a cloned DNA sequence".

The term "isolated", when applied to a protein, refers to a protein that is found in a condition other than its native environment. In a preferred form, the isolated protein is substantially free of other proteins, particularly other homologous proteins. An isolated protein is more than 10% pure, preferably more than 20% pure, more preferably more than 30% pure, as determined by SDS-PAGE. Further aspects of the invention encompass the protein in a highly purified form, i.e., more than 40% pure, more than 60% pure, more than 80% pure, more than 90% pure, more than 95% pure, more than 97% pure, and even more than 99% pure, as determined by SDS-PAGE.

The term "oxidation stable" refers to a *Cellulomonas* protease that retains a specified amount of enzymatic activity over a given period of time under conditions prevailing during the proteolytic, hydrolyzing, cleaning or other process of the invention, for example while exposed to or contacted with bleaching agents or oxidizing agents. In one aspect the serine protease retains at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% proteolytic activity after contact with a bleaching or oxidizing agent over a given time period, for example, at least 1 minute, 3 minutes, 5 minutes, 8 minutes, 12 minutes, 16

minutes, 20 minutes, etc.. The stability may, e.g., be measured as described in the section entitled "Determination of Oxidative stability" hereinafter.

The term "chelator stable" refers to a *Cellulomonas* protease that retains a specified amount of enzymatic activity over a given period of time under conditions prevailing during the proteolytic, hydrolyzing, cleaning or other process of the invention, for example while exposed to or contacted with chelating agents. In one aspect the serine protease retains at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% proteolytic activity after contact with a chelating agent over a given time period, for example, at least 10 minutes, 20 minutes, 40 minutes, 60 minutes, 100 minutes, etc. The stability may, e.g., be measured as described in the section entitled "Determination of chelate stability" hereinafter.

The term "thermally stable" refers to a serine protease that retains a specified amount of enzymatic activity after exposure to identified temperatures over a given period of time under conditions prevailing during the proteolytic, hydrolyzing, cleaning or other process of the invention, for example while exposed altered temperatures. Altered temperatures includes increased or decreased temperatures. In one aspect the serine protease retains at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% proteolytic activity after exposure to altered temperatures over a given time period, for example, at least 60 minutes, 120 minutes, 180 minutes, 240 minutes, 300 minutes, etc. The stability may, e.g., be measured as described in the section entitled "Determination of thermal stability" hereinafter.

The term "enhanced stability" in the context of an oxidation, chelator, thermal and/or pH stable protease refers to a higher retained proteolytic activity over time as compared to other serine proteases, e.g., subtilisin proteases.

The term "diminished stability" in the context of an oxidation, chelator, thermal and/or pH stable protease refers to a lower retained proteolytic activity over time as compared to other serine proteases, e.g., subtilisin proteases.

The term "cleaning activity" refers to the cleaning performance achieved by the protease under conditions prevailing during the proteolytic, hydrolyzing, cleaning or other process of the invention. Cleaning performance can be determined by the application of various cleaning assays concerning enzyme sensitive stains, for example grass, blood, milk, or egg protein as determined by various chromatographic, spectrophotometric or other quantitative methodologies after subjection of the stains to standard wash conditions. Exemplary assays include those described in WO 99/34011, US Patent 6,605,458 and those described in the section entitled "Determination of Cleaning Activity" hereinafter.

The term "cleaning effective amount" of a protease refers to the quantity of protease described hereinbefore that achieves a desired level of enzymatic activity in a specific cleaning composition. Such effective amounts are readily ascertained by one of ordinary skill in the art and is based on many factors, such as the particular protease used, the cleaning application, the specific composition of the cleaning composition, and whether a liquid or dry (e.g., granular, bar) composition is required, and the like.

The term "cleaning adjunct materials", as used herein, means any liquid, solid or gaseous material selected for the particular type of cleaning composition desired and the form of the product (e.g., liquid; granule; powder; bar; paste; spray; tablet; gel; foam composition), which materials are also preferably compatible with the protease enzyme used in the composition. Granular compositions can also be in "compact" form and the liquid compositions can also be in a "concentrated" form.

The term "enhanced performance" in the context of cleaning activity refers to an increased or greater cleaning activity of certain enzyme sensitive stains such as egg, milk, grass or blood, as determined by usual evaluation after a standard wash cycle.

The term "diminished performance" in the context of cleaning activity refers to an decreased or lesser cleaning activity of certain enzyme sensitive stains such as egg, milk, grass or blood, as determined by usual evaluation after a standard wash cycle.

The term "comparative performance" in the context of cleaning activity refers to at least 60%, at least 70%, at least 80% at least 90% at least 95% of the cleaning activity of a comparative subtilisin protease, e.g., wild-type *Bacillus licheniformis* proteases (for example OPTIMASE brand protease products sold by Genencor International, Palo Alto, CA, USA), wild-type GG36 proteases (for example PURAFECT ™ protease products by Genencor International, Inc, Palo Alto, CA, USA and/or SAVINASE ™ protease products sold by Novozymes), BPN'-variants (for example those described in US RE 34,606), GG36-variants (for example, RELASE™, DURAZYME™, EVERLASE™, KANNASE ™ protease products sold by Novozymes; and/or MAXACAL™, MAXAPEM™, PROPERASE ™ protease products sold by Genencor International; and/or those described in US Patent RE 34,606, US Patent Nos. 5,700,676; 5,955,340; 6,312,936; 6,482,628; *Bacillus lentus* variant protease products [for example those described in WO 92/21760, WO 95/23221 and/or WO 97/07770 sold by Henkel KgaA, Dusselfordf, Germany). Exemplary subtilisin protease variants include those having substitutions or deletions at residue positions equivalent to positions 76, 101, 103, 104, 120, 159, 167, 170, 194, 195, 232, 235, 236, 245, 248, and/or 252 of BPN'. Cleaning

performance can be determined by comparing the proteases of the present invention with those subtilisin proteases in various cleaning assays concerning enzyme sensitive stains such as grass, blood or milk as determined by usual spectrophotometric or analytical methodologies after standard wash cycle conditions.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of protein purification, molecular biology, microbiology, recombinant DNA techniques and protein sequencing, which are within the skill of the art. Such techniques are explained fully in the literature. See, e.g., A Practical Guide to protein and Purification for Microsequencing, (P. T. Matsudaira, ed., (1989) Academic Press, Inc.); Sambrook, Fritsch & Maniatis, Molecular Cloning; A Laboratory Manual; Second Edition (1989); Oligonucleotide Synthesis (M. J. Gait, ed., 1984); Nucleic Acid Hybridization (B. D. Hames & S. J. Higgins, eds., 1984); A Practical Guide to Molecular Cloning (B. Perbal, 1984); PCR Protocols, A Guide to Methods and Applications (M. A. Innis, et al, eds., (1990) Academic Press, Inc.); Current Protocols in Molecular Biology (F. M. Ausubel, et al, eds., (1989) John Wiley & Sons); Laboratory Techniques in Biochemistry and Molecular Biology, Vol. 9, G. Allen, Second Edition (1989) Elsevier; and additional publications in the series, Methods in Enzymology (Academic Press, Inc.). All patents, patent applications, and publications mentioned herein, both supra and infra, are hereby incorporated by reference.

II. Serine protease enzymes and nucleic acid encoding serine protease enzymes

One aspect of the invention encompasses isolated polynucleotides encoding an amino acid sequence, e.g., encoding a protease, comprising at least 65% amino acid sequence identity, at least 70% amino acid sequence identity, at least 75% amino acid sequence identity, at least 80% amino acid sequence identity, at least 85% amino acid sequence identity, at least 90% amino acid sequence identity, at least 92% amino acid sequence identity, at least 95% amino acid sequence identity, at least 97% amino acid sequence identity, at least 98% amino acid sequence identity and at least 99% amino acid sequence identity to the amino acid sequence as shown in Figure 8 (SEQ ID. NOS.:8), the amino acid sequence, e.g., protease, encoded by the polynucleotide having proteolytic activity, e.g., catalyzing the hydrolysis of peptide linkages of substrates and/or demonstrate comparable or enhanced washing performance under identified wash conditions.

One aspect of the invention encompasses polynucleotides encoding an amino acid sequence having at least 65% amino acid sequence identity, at least 70% amino acid sequence identity, at least 75% amino acid sequence identity, at least 80% amino acid

sequence identity, at least 85% amino acid sequence identity, at least 90% amino acid sequence identity, at least 92% amino acid sequence identity, at least 95% amino acid sequence identity, at least 97% amino acid sequence identity, at least 98% amino acid sequence identity and at least 99% amino acid sequence identity to the amino acid sequence as shown in Figure 6, 7, and/or 8 (SEQ ID. NO.:6, 7, 8), e.g., a portion of the amino acid sequence encoded by the polynucleotide having proteolytic activity, e.g., the mature protease catalyzing the hydrolysis of peptide linkages of substrates and/or demonstrate comparable or enhanced washing performance under identified wash conditions.

Percent identity (amino acid sequence, nucleic acid sequence, gene sequence) can be determined, for example, by a direct comparison of the sequence information between two molecules by aligning the sequences, counting the exact number of matches between the two aligned sequences, dividing by the length of the shorter sequence, and multiplying the result by 100. Readily available computer programs can be used to aid in the analysis, such as ALIGN, Dayhoff, M.O. in Atlas of Protein Sequence and Structure M.O. Dayhoff ed.,* 5 Suppl. 3:353-358, National biomedical Research Foundation, Washington, DC, which adapts the local homology algorithm of Smith and Waterman (1981) Advances in Appl. Math. 2:482-489 for peptide analysis. Programs for determining nucleotide sequence identity are available in the Wisconsin Sequence Analysis Package, Version 8 (available from Genetics Computer Group, Madison, WI) for example, the BESTFIT, FASTA and GAP programs, which also rely on the Smith and Waterman algorithm. These programs are readily utilized with the default parameters recommended by the manufacturer and described in the Wisconsin Sequence Analysis Package referred to above.

An example of an algorithm that is suitable for determining sequence similarity is the BLAST algorithm, which is described in Altschul, *et al., J. Mol. Biol. 215*:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. These initial neighborhood word hits act as starting points to find longer HSPs containing them. The word hits are expanded in both directions along each of the two sequences being compared for as far as the cumulative alignment score can be increased. Extension of the word hits is stopped when: the cumulative alignment score falls off by the quantity X from a maximum achieved

value; the cumulative score goes to zero or below; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a wordlength (W) of 11, the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA 89*:10915 (1989)) alignments (B) of 50, expectation (E) of 10, M'5, N'-4, and a comparison of both strands.

The BLAST algorithm then performs a statistical analysis of the similarity between two sequences (*see, e.g.,* Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA 90*:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a serine protease nucleic acid of this invention if the smallest sum probability in a comparison of the test nucleic acid to a serine protease nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001. Where the test nucleic acid encodes a serine protease polypeptide, it is considered similar to a specified serine protease nucleic acid if the comparison results in a smallest sum probability of less than about 0.5, and more preferably less than about 0.2.

In one aspect of the invention, sequences were analyzed by BLAST and protein translation sequence tools. In one aspect the preferred version is BLAST (Basic BLAST version 2.0). The program chosen was "blastx", and the database chosen was "nr". Standard/default parameter values were employed.

One aspect of the invention encompasses a polynucleotide shown in Figs.1A-B, having SEQ. ID NO: 1. approximately 1621 base pairs in length. An initiation peptide is bolded in Fig. 1. In another aspect of the present invention, the polynucleotides encoding these amino acid sequences comprise a 1485 base pair portion (residues 1-1485 of Figs 2A-B) that, if expressed, is believed to encode a signal sequence (nucleotides 1-84 of Fig. 5 [SEQ ID. No. 5] encoding amino acids 1-28 of Fig 9 [SEQ ID. NO.9]), an N-terminal prosequence (nucleotides 84-594 encoding amino acid residues 29-198 of Fig. 6), a mature protease sequence (nucleotides 595-1161 of Figure 2B [SEQ ID NO. 2] encoding amino acid residues 1-189 of Fig. 8 [SEQ ID NO. 8]) and a C-terminal pro-sequence (nucleotides 1346-1672 encoding amino acid residues 388-495 of Fig. 6[SEQ ID NO. 6]). Alternatively, the signal peptide, the N-terminal pro-sequence, mature serine protease sequence and C-terminal pro-sequence can be numbered in relation to the amino acid residues of the mature protease of Figure 6 being numbered 1-189, i.e., signal peptide (residues -198 to –171), an N-terminal pro sequence (residues -171 to -1), the mature serine protease sequence

(residues 1-189) and a C-terminal pro-sequence (residues 190-298). In another aspect of the present invention, the polynucleotide encoding an amino acid sequence having proteolytic activity, e.g., a protease, comprises a nucleotide sequence of nucleotides 1 to 1485 of Figs 2A-B (the portion of SEQ ID NO.2 encoding the signal peptide and precursor protease). In another aspect of the invention, the polynucleotide encoding an amino acid sequence comprises the sequence of nucleotides 1 to 1412 of Figs 3A-B (the polynucleotide encoding the precursor *Cellulomonas* protease, e.g., SEQ ID NO. 3). In another aspect of the invention, the polynucleotide encoding an amino acid sequence comprises the sequence of nucleotides 1 to 587 of Fig 4 (the portion of the polynucleotide encoding the mature *Cellulomonas* protease, e.g., SEQ ID NO. 4).

As will be understood by the skilled artisan, due to the degeneracy of the genetic code, a variety of polynucleotides can encode the signal peptide, precursor protease and/or mature protease disclosed in Figures 6, 7, and/or 8 (SEQ ID NOS.: 6, 7, and/or 8, respectively) or a protease having the % sequence identity described above. Another aspect of the invention encompasses a polynucleotide comprising a nucleotide sequence having at least 70% sequence identity, at least 75% sequence identity, at least 80% sequence identity, at least 85% sequence identity, at least 90% sequence identity, at least 92% sequence identity, at least 95% sequence identity, at least 97% sequence identity, at least 98% sequence identity and at least 99% sequence identity to the polynucleotide sequence of Figs. 2, 3, and/or 4 (SEQ ID. NOS. 2, 3, and/or 4, respectively), encoding the signal peptide and precursor protease, the precursor protease and/or the mature protease, respectively.

Another aspect of the invention includes fragments or portions of such DNA fragments encoding proteases, so long as the encoded fragment retains proteolytic activity. Another aspect of the invention encompasses polynucleotides having at least 20% of the sequence length, at least 30% of the sequence length, at least 40% of the sequence length, at least 50% of the sequence length, at least 60% of the sequence length, 70% of the sequence length, at least 75% of the sequence length, at least 80% of the sequence length, at least 85% of the sequence length, at least 90% of the sequence length, at least 92% of the sequence length, at least 95% of the sequence length, at least 97% of the sequence length, at least 98% of the sequence length and at least 99% of the sequence of the polynucleotide sequence of Fig.2 (SEQ ID. NO. 2) or residues 185-1672 of Fig. 1 (SEQ. ID. NO.:1), encoding the precursor protease. In one aspect of the invention, such fragments or portions of the sequence length are contiguous portions of the sequence length, useful for

shuffling of the DNA sequence in recombinant DNA sequences (US Patent No. 6,132,970; Stemmer, WP; October 17, 2000)

Another aspect of the invention includes fragments of the DNA described herein that may be used according to art recognized techniques for the purpose of obtaining partial length DNA fragments capable of being used to isolate or identify polynucleotides encoding mature protease enzyme described herein from *Cellulomonas* 69B4, or a segment thereof having proteolytic activity. Moreover, the DNA provided in Figs.1A-B will be useful for obtaining homologous fragments of DNA from other species, and particularly from *Cellulomonas spp.* which encode a protease or portion thereof having proteolytic activity.

In addition, the present invention encompasses using primer or probe sequences constructed from SEQ ID NO:1, or a suitable portion or fragment thereof (e.g.; at least about 5-20 or 10-15 contiguous nucleotides), as a probe or primer for screening nucleic acid of either genomic or cDNA origin. A DNA probe of the desired length, in one aspect generally between 100 and 1000 bases in length, taken from the sequences in FIGS. 1, 2, 3 or 4 (SEQ ID NOS.: 1, 2, 3, and/or 4, respectively) should be isolated by electrophoresis in an agarose gel, the fragment excised from the gel, and recovered from the excised agarose. For a more detailed procedure, see Sambrook. This purified fragment of DNA is then labeled (using, for example, the Megaprime labeling system according to the instructions of the manufacturer) to incorporate P^{32} in the DNA. The labeled probe is denatured by heating to 95.degree. C. for a given period of time, e.g., 5 minutes, and immediately added to the membrane and prehybridization solution. The hybridization reaction should proceed for an appropriate time and under appropriate conditions, for example, for 18 hours at 37 ° C. with gentle shaking or rotating. The membrane is rinsed (for example, in 2.times.SSC/0.3% SDS) and then washed in an appropriate wash solution with gentle agitation. The stringency desired will be a reflection of the conditions under which the membrane (filter) is washed. "Low-stringency" conditions, in this context, refers to washing with a solution of 0.2X SSC/0.1% SDS at 20° C for 15 minutes. "Medium-stringency" conditions, in this context, refers to a further washing step comprising washing with a solution of 0.2X SSC/0.1% SDS at 37°C for 30 minutes. "High-stringency" conditions in this context refers to a further washing step comprising washing with a solution of 0.2X SSC/0.1% SDS at 37°C for 45 minutes. "Maximum-stringency" conditions,in this context, refers to a further washing step comprising washing with a solution of 0.2X SSC/0.1% SDS at 37°C for 60 minutes. Thus aspects of the present invention includes those polynucleotides capable of hybridizing to a probed derived from the nucleotide sequence disclosed in Figures 1, 2, 3, 4 and/or 5 (SEQ

ID Nos.:1, 2, 3, 4, and/or 5) under conditions of medium, high and/or maximum stringency.

After washing, the membrane is dried and the bound probe detected. If P^{32} or another radioisotope is used as the labeling agent, the bound probe can be detected by autoradiography. Other techniques for the visualization of other probes are well-known to those of skill. The detection of a bound probe indicates a nucleic acid sequence has the desired homology, and therefore identity to SEQ ID NOS:1, 2, 3, 4, and/or 5, and is encompassed within this invention. Accordingly, the present invention provides a method for the detection of nucleic acid encoding a protease encompassed by the present invention which comprises hybridizing part or all of a nucleic acid sequence of SEQ ID NOS:1, 2, 3, 4, and / or 5 with *Cellulomonas* nucleic acid of either genomic or cDNA origin.

Another aspect of the invention included within the scope of the present invention is the use of hybridization conditions based on the melting temperature (Tm) of the nucleic acid binding complex, as taught in Berger and Kimmel (1987, Guide to Molecular Cloning Techniques, Methods in Enzymology, Vol. 152, Academic Press, San Diego, Calif. incorporated herein by reference, to confer a defined "stringency" as explained below. "Maximum stringency" typically occurs at about Tm-5° C. (5° C. below the Tm of the probe); "high stringency" at about 5° C. to 10° C. below Tm; "intermediate stringency" at about 10° C. to 20° C. below Tm; and "low stringency" at about 20° C. to 25° C. below Tm. As will be understood by those of skill in the art, medium, high and /or maximum stringency hybridization can be used to identify or detect polynucleotide sequence homologues or equivalent polynucleotide sequences.

Another aspect of the invention encompasses a nucleic acid construct comprising the aforementioned polynucleotide. Another aspect of the invention encompasses a vector comprising the aforementioned polynucleotide. Another aspect of the invention encompasses a host cell transformed with the aforedescribed vector.

Another aspect of the invention encompasses those polynucleotides encoding a signal sequence. In one aspect of the invention encompasses polynucleotides having signal activity comprising a nucleotide sequence having at least 65% sequence identity, at least 70% sequence identity, at least 75% sequence identity, at least 80% sequence identity, at least 85% sequence identity, at least 90% sequence identity, at least 95% sequence identity, at least 97% sequence identity, at least 98% sequence identity and at least 99% sequence identity to SEQ ID NO.:5, a putative signal sequence, polynucleotides being capable of hybridizing to a probed derived from the nucleotide sequence disclosed in SEQ ID NO.:5 under conditions of medium, high and or maximal stringency, said signal

sequences having substantially the same signal activity as said signal sequence encoded by the polynucleotide of the present invention.

Substantially the same signal activity can be indicated by substantially the same secretion of the protease into the fermentation medium, for example a fermentation medium protease level being at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98% of the secreted protease levels in the fermentation medium as provided by the signal sequence of SEQ ID No. 3. The secreted protease levels can be ascertained by protease activity studies, e.g., pNA assay (DelMar, 1979). Additional means for determining the levels of secretion of a heterologous or homologous protein in a gram positive host cell and detecting secreted proteins include using either polyclonal or monoclonal antibodies specific for the protein. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS). These and other assays are described, among other places, in Hampton, R. et al., (1990, Serological Methods, a Laboratory Manual, APS Press, St. Paul Minn.) and Maddox, D E et al., (1983, J. Exp. Med., 158:1211).

Another aspect of the invention encompasses those polynucleotides (nucleotides 1-84 of Fig. 5 (SEQ ID NO.:5) encoding an amino acid sequence of a signal peptide as shown in Fig. 9 (SEQ ID NO.:9), nucleotide residue positions 1to 85 of Fig 2 (SEQ ID NO.2) and /or SEQ ID NO. 5. The invention further encompasses nucleic acid sequences which hybridize to the nucleic acid sequence shown in SEQ ID NO:5 under the following conditions, under low, under medium, under high stringency conditions, and/or maximum stringency conditions, but which have substantially the same signal activity as said sequence. The present invention encompasses all such polynucleotides.

Another aspect of the invention encompasses polynucleotides being complementary to the nucleotide sequences described herein. Exemplary complementary nucleotide sequences are shown in Figs 1-5 (SEQ ID NOS.1-5, respectively).

Further aspects of the present invention encompass polypeptides having proteolytic activity, e.g., proteases, comprising 65% amino acid sequence identity, at least 70% sequence identity, at least 75% amino acid sequence identity, at least 80% amino acid sequence identity, at least 85% amino acid sequence identity, at least 90% amino acid sequence identity, at least 92% amino acid sequence identity, at least 95% amino acid sequence identity, at least 97% amino acid sequence identity, at least 98% amino acid sequence identity and at least 99% amino acid sequence identity to the amino acid sequence of Figure 6 (the signal and precursor protease) (SEQ. ID. NO. 6), Figure 7 (the

precursor protease) (SEQ ID NO. 7) and/or of Figure 8 (the mature protease) (SEQ ID NO.:8), the polypeptide, e.g., protease, having proteolytic activity, e.g., catalyzing the hydrolysis of peptide linkages of substrates or demonstrate comparable or enhanced washing performance under identified wash conditions. In another aspect of the present invention, the polypeptides are isolated. In another aspect of the present invention, the polypeptides comprise an amino acid sequence identical to the amino acid sequence of Figure 6, Figure 7 or of Figure 8.

One aspect of the invention encompasses an isolated polypeptide having proteolytic activity, e.g., a protease, comprising the amino acid sequence shown in Fig 6 (SEQ. ID NO: 6), approximately 495 amino acids in length. Another aspect of the present invention encompasses a polypeptide having proteolytic activity comprising the amino acid sequence shown in Fig. 7 (SEQ ID NO.:7), approximately 467 amino acids in length.. In one aspect, these amino acid sequences comprise a signal sequence (amino acids 1-28 of Fig. 9 [SEQ ID NO.:9]);and a precursor protease (amino acids 1-467 of Fig. 7 [SEQ ID NO.:7]). Another aspect of the present invention encompasses a polypeptide comprising an N-terminal prosequence (amino acids 1-170 of Fig. 7 [SEQ ID NO.:7]), a mature protease sequence (amino acids 1-189 of Figure 8 [SEQ ID NO. 8]), and a C-terminal prosequence (amino acids 360 –467 of Fig 7[SEQ ID NO. 7]). Another aspect of the present invention encompasses a polypeptide comprising a precursor protease sequence (amino acids 1-467 of Fig. 7 [SEQ ID NO.:7]). Another aspect of the present invention encompasses a mature protease sequence comprising amino acids 1-189 of Figure 8 [SEQ ID NO. 8])

The present invention encompasses polypeptides and/or proteases comprising amino acid sequences of the above described % sequence identities derived from bacterial species including, but not limited to *Cellulomonas* spp which can be identified through amino acid sequence homology studies. A residue (amino acid) of a precursor *Cellulomonas* protease is equivalent to a residue of *Cellulomonas* strain 69B4 if it is either homologous (i.e., corresponding in position in either primary or tertiary structure) or analogous to a specific residue or portion of that residue in *Cellulomonas* strain 69B4 protease (i.e., having the same or similar functional capacity to combine, react, or interact chemically).

In order to establish homology to primary structure, the amino acid sequence of a precursor protease is directly compared to the *Cellulomonas strain 69B4* mature protease amino acid sequence and particularly to a set of conserved residues which are discerned to be invariant in all or a large majority of *Cellulomonas* like proteases for which sequence is

known. After aligning the conserved residues, allowing for necessary insertions and deletions in order to maintain alignment (i.e., avoiding the elimination of conserved residues through arbitrary deletion and insertion), the residues corresponding to particular amino acids in the mature protease (Fig. 8, SEQ ID NO.:8) and *Cellulomonas* 69B4 protease can be determined. Alignment of conserved residues preferably should conserve 100% of such residues. However, alignment of greater than 75% or as little as 50% of conserved residues is also adequate to define equivalent residues. Conservation of the catalytic triad, His32/Asp56/Ser137 of Fig 8 (SEQ ID.NO.: 8) should be maintained.

For example, the amino acid sequence of proteases from *Cellulomonas* strain 69B4, and *Cellulomonas* spp. described supra may be aligned to provide the maximum amount of homology between amino acid sequences. A comparison of these sequences would show that there are a number of conserved residues contained in each sequence. These are the residues that would be identified and utilized to establish the equivalent residue positions of amino acids identified in the precursor or mature *Cellulomonas* protease in question.

These conserved residues thus may be used to ascertain the corresponding amino acid residues of *Cellulomonas* strain 69B4 protease in one or more in *Cellulomonas* homologues, e.g., *Cellulomonas cellasea* (DSM 20118) and/or a *Cellulomonas* homologue herein. These particular amino acid sequences can be aligned with the sequence of *Cellulomonas* 69B4 protease to produce the maximum homology of conserved residues. By this alignment, the sequences and particular residue positions of *Cellulomonas* 69B4 can be observed as compared to other *Cellulomonas* spp. Thus, the equivalent amino acid for the catalytic triad, e.g., in *Cellulomonas* 69B4 protease can be identified in the other *Cellulomonas* spp.. In one aspect of the present invention, the protease or *Cellulomonas* homolog comprises the equivalent of His32/Asp56/Ser137 of Fig 8 (SEQ ID.NO.: 8)

Another indication that two polypeptides are substantially identical is that the first polypeptide is immunologically cross-reactive with the second polypeptide. Methodologies for determining immunological cross-reactivity are described in the art. Typically, polypeptides that differ by conservative amino acid substitutions are immunologically cross-reactive. Thus, a polypeptide is substantially identical to a second polypeptide, for example, where the two peptides differ only by a conservative substitution.

Aspects of the present invention encompass the aforementioned polynucleotides and polypeptides derived from a bacterial or fungal source. In another aspect, the polynucleotides and/or polypeptides are isolated from a bacterial or fungal source. In one

aspect the bacterial source is a *Cellulomonas* spp. In another aspect of the present invention, a *Cellulomonas* spp include or can be selected from the group of *C. fimi* (ATCC 484, DSM 20113), *C. biazotea* (ATCC 486, DSM 20112), *C. cellasea* (ATCC 487, 21681, DSM 20118), *C. hominis* (DSM 9581), C *flavigena* (DSM 20109), *C. persica* (ATCC 700642, DSM 14784), *C. iranensis* (ATCC 700643, DSM 14785); *C. fermentans* (DSM 3133), *C. gelida* (ATCC 488, DSM 20111), *C. humilata* (ATCC 25174), *C. turbata* [re-evaluated as *Oerskovia turbata*] (ATCC 25835, DSM 20577), *C. uda* (ATCC 491, DSM 20107) and *C. strain 69B4* (DSM 16035). In another aspect the bacterial source may also include or can be selected from the group of *Thermobifida* spp., *Rarobacter* spp., and/or *Lysobacter* spp. In another aspect, the *Thermobifida* spp. is *Thermobifida fusca* (tfpA, AAC23545, Lao, G., et. al, Appl. Environ. Microbiol. 62(11):4256-4259 (1996)). In another aspect, the *Rarobacter* spp. is *Rarobacter faecitabidus* (RPI, A45053, Shimoi, H.; et al, J. Biol. Chem. 267:25189-25195 (1992)). In another aspect the *Lysobacter* spp. is *Lysobacter enzymogenes*.

In another aspect of the present invention the polypeptides and/or polynucleotides are derived from a fungal source. In another aspect the polypeptides and/or polynucleotides are isolated from a fungal source. In another aspect the fungal source includes a *Metarhizium* spp. In another aspect, the fungal source is a *Metarhizium anisopliae* (CHY1 (CAB60729).

In another aspect of the present invention, the polypeptide and/or polynucleotides are derived from a *Cellulomonas* strain selected from cluster 2 of the taxonomic classification described in US Patent No 5,401,657.

In US Patent 5,401,657, twenty strains of bacteria isolated from in and around alkaline lakes were assigned to the type of bacteria known as Gram-positive bacteria on the basis of (1) the Dussault modification of the Gram's staining reaction (Dussault, H.P., (1955), Journal of Bacteriology, 70, 484-485); (2) the KOH sensitivity test (Gregersen, T., (1978), European Journal of Applied Microbiology and Biotechnology 5, 123-127; Halebian, S. et al., (1981), Journal of Clinical Microbiology, 13, 444-448); (3) the aminopeptidase reaction (Cerny, G., (1976), European Journal of Applied Microbiology, 3, 223-225; ibid, (1978), 5, 113-122); and in most cases, confirmation also on the basis of (4) a quinone analysis (Collins, M.D. and Jones, D., (1981), Microbiological Reviews, 45, 316-354) using the method described by Collins, M.D. in Chemical Methods in Bacterial Systematics (eds. Goodfellow, M. and Minnikin, D.) pp. 267-288, Academic Press, London, 1985. Strains can be tested for 200 characters and the results analyzed using the principles of numerical

taxonomy (Sneath, P.H.A. and Sokal, R.R., in Numerical Taxonomy, W.H. Freeman & Co.,. San Francisco, 1973). Exemplary characters tested and manner of testing are described in US Patent 5,401,657, Appendix B. In addition, US Patent 5,401,657, Appendix C records how each character was coded for taxonomic analysis.

The phenetic data, consisting of 200 unit characters was scored as indicated in Appendix C, and set out in the form of an "n.times.t" matrix, whose t columns represent the t bacterial strains to be grouped on the basis of resemblances, and whose n rows are the unit characters. Taxonomic resemblance of the bacterial strains was estimated by means of a similarity coefficient (Sneath, P.H.A. and Sokal, R.R., Numerical Taxonomy, supra, pp. 114-187). Although many different coefficients have been used for biological classification, only a few have found regular use in bacteriology. Three association coefficients (Sneath, P.H.A. and Sokal, R.R., ibid, p. 129 et seq.), namely, the Gower, Jaccard and Simple Matching coefficients were applied. These have been frequently applied to the analysis of bacteriological data and have a wide acceptance by those skilled in the art since they have been shown to result in robust classifications.

The coded data were analyzed using the TAXPAK program package (Sackin, M.J., "Programmes for classification and identification". In Methods in Microbiology, Volume 19 (eds. R.R. Colwell and R. Grigorova), pp. 459-494, Academic Press, London, (1987)) run on a DEC VAX computer at the University of Leicester, U.K.

A similarity matrix was constructed for all pairs of strains using the Gower Coefficient (S_G) with the option of permitting negative matches (Sneath, P.H.A. and Sokal, R.R., supra, pp. 135-136) using the RTBNSIM program in TAXPAK. As the primary instrument of analysis and the one upon which most of the arguments presented herein are based, the Gower Coefficient was chosen over other coefficients for generating similarity matrices because it is applicable to all types of characters or data, namely, two-state, multistate (ordered and qualitative), and quantitative.

Cluster analysis of the similarity matrix was accomplished using the Unweighted Pair Group Method with Arithmetic Averages (UPGMA) algorithm, also known as the Unweighted Average Linkage procedure, by running the SMATCLST sub-routine in TAXPAK.

A dendrogram, would illustrate the levels of similarity between the bacterial strains. The dendrogram is obtained by using the DENDGR program in TAXPAK. The phenetic data, were re-analyzed using the Jaccard Coefficient (S_J) (Sneath, P.H.A. and Sokal, R.R., ibid, p. 131) and Simple Matching Coefficient (S_{SM}) (Sneath, P.H.A. and Sokal, R.R., ibid, p. 132) by running the RTBNSIM program in TAXPAK. A further two dendrograms were

obtained by using the SMATCLST with UPGMA option and DENDGR sub-routines in TAXPAK.

S_G /UPGMA Method

Six natural clusters or phenons of alkaliphilic bacteria are generated at the 79% similarity level. These six clusters include 15 of the 20 alkaliphilic bacteria isolated from alkaline lakes. Although the choice of 79% for the level of delineation may seem arbitrary, it is in keeping with current practices in numerical taxonomy (Austin, B. and Priest, F., in Modern Bacterial Taxonomy, p. 37; Van Nostrand Reinhold; Wokingham, U.K., (1986)). Placing the delineation at a lower percentage would combine groups of clearly unrelated organisms whose definition is not supported by the data. At the 79% level, 3 of the clusters exclusively contain novel alkaliphilic bacteria representing 13 of the newly isolated strains, and these may represent new taxa. Protease 69B4 was classified as in cluster 2 by this method.

The significance of the clustering at this level was supported by the results of the TESTDEN program. This program tests the significance of all dichotomous pairs of clusters (comprising 4 or more strains) in a UPGMA generated dendrogram with Squared Euclidean distances, or their complement as a measurement and assuming that the clusters are hyperspherical. The critical overlap was set at 0.25%. As can be seen from Table 5, the separation of the clusters is highly significant.

S_J /UPGMA and S_{SM} /UPGMA Methods

The S_J coefficient is a useful adjunct to the S_G coefficient as it can be used to detect phenons in the latter that are based on negative matches or distortions owing to undue weight being put on potentially subjective qualitative data. Consequently, the S_J coefficient is useful for confirming the validity of clusters defined initially by the use of the S_G coefficient. The Jaccard Coefficient is particularly useful in comparing biochemically unreactive organisms (Austin, B. and Priest, F.G., supra, p. 37). There may be doubts about the admissability of matching negative character states (Sneath, P.H.A., and Sokal, R.R., supra, p. 131) in which case the Simple Matching Coefficient is a widely applied alternative. Strain 69B4 was classified as in cluster 2 by this method.

In the main, all of the clusters (especially the clusters of the new bacteria) generated by the S_G /UPGMA method are recovered in the dendrograms produced by the S_J /UPGMA method (cophenetic correlation, 0.795) and the S_{SM} /UPGMA method (cophenetic correlation, 0,814). The main effect of these transformations is to gather all the *Bacillus* strains in a single large cluster which further serves to emphasize the separation between

the alkaliphilic *Bacillus* species and the new alkaliphilic bacteria, and the uniqueness of the latter. Based on these methodologies, 69B4 is considered to be a cluster 2 bacterium.

In other aspects of the present invention, the polynucleotide is derived from a bacteria having a 16S rRNA gene nucleotide sequence at least 70%, 75%, 80%, 85%, 88%, 90%, 92%, 95%, 98% sequence identity with the 16S rRNA gene nucleotide sequence of *Cellulomonas* strain 69B4. The sequence of the 16S rRNA gene is deposited at GenBank under Accession Number X92152 (Lake Bogoria isolate 69B4). A phylogenetic tree can be constructed from aligned sequences using TREECON for Windows computer program (van de Peer, Y., University of Antwerp, Antwerpen, Belgium) (van de Peer and de Wachter, Comput Appl. Biosci. 10(5): 569-70 (1994)). Distance estimations were calculated using the substitution rate calibration of Jukes and Cantor (Jukes, T. H., et al, "Evolution of protein molecules", pp 21-32 in Mammalian Protein Metabolism (Munro, H. N.; ed.) Academic Press, New York (1969)) and can be utilized to determine the % sequence identity. The tree topology was inferred by the Neighbor-Joining algorithm and rooted to the *E. coli* sequence. The numbers at the nodes refer to bootstrap values (100 resamplings). Bar = 2% sequence divergence. The resulting phylogenetic tree for *Cellulomonas* spp., including *Cellulomonas* strain 69B4 is shown in Fig. 10. The strain 69B4 exhibits the closest 16S rDNA relationship to members of *Cellulomonas* and *Oerskovia* of the family *Cellulomonadaceae*. The closest relatives are believed to be *C. cellasea* (DSM 20118) and *C. fimi* (DSM 20113) with at least 95% sequence identity with the 16S rRNA gene nucleotide sequence of *Cellulomonas* strain 69B4 (e.g., 96% and 95% identity respectively) to strain 69B4 16S rRNA gene sequence.

In another aspect of the invention, the *Cellulomonas* spp. is *Cellulomonas* strain 69B4 (DSM16035). Strain 69B4 was originally isolated from sediment and water from the littoral zone of Lake Bogoria, Kenya at Acacia Camp (Lat. 0° 12'N, Long. 36° 07'E) collected on 10 October 1988. The water temperature was 33°C, pH 10.5 with a conductivity of 44 mS/cm. *Cellulomonas* strain 69B4 was deposited on November 12, 2003, in accordance with the Budapest treaty at the Deutsche Sammlung von Mikroorganismen und Zellkulturen Gmbh ("DSMZ") [German Collection of Microorganisms and Cell Cultures] at Mascheroder Weg 1b, D-38124 Braunschweig, Germany, Deposit Number DSM 16035. Strain 69B4 was isolated at 37° C on an alkaline casein medium as described in US Patent 5,401,657 (issued March 28, 1995) and in Duckworth, A.W., et al , FEMS Microbiology Ecology, 19:191 (1996).

Cellulomonas strain 69B4 had the following phenotypic characteristics. Fresh cultures were, Gram-positive, slender, generally straight, rod-shaped bacterium, approximately 0.5-0.7μm x 1.8-4μm. Older cultures contained mainly short rods and

coccoid cells. Cells occasionally occurred in pairs or as V-forms, but primary branching was not observed. Endospores were not detected. On alkaline GAM agar the strain forms opaque, glistening, pale-yellow coloured, circular and convex or domed colonies, with entire margins, about 2 mm in diameter after 2-3 days incubation at 37°C. The colonies were viscous or slimy with a tendency to clump when scraped with a loop. On neutral Tryptone Soya Agar, strain growth was less vigorous giving translucent yellow colonies, generally <1 mm in diameter. The cultures were aerobic and also facultative anaerobe capable of growth under strictly anaerobic conditions. Growth under anaerobic conditions was markedly reduced compared to aerobic growth.

The strain also appeared to be negative for oxidase, urease, aminopeptidase and KOH tests. Nitrate was not reduced. There was a positive catalase reaction. Dnase was produced under alkaline conditions.

The temperature range for growth was 20 - 37°C with an optimum around 30-37°C. No growth was observed at 15°C or 45°C.

The strain can also be characterized as being alkaliphilic and slightly halophilic. The strain may also be characterized as having growth occurring at pH values between 6.0 and 10.5 with an optimum around pH 9-10. No growth was observed at pH 11 or pH 5.5. Growth below pH7 was less vigorous and abundant. The strain was observed to grow in medium containing 0-8% (w/v) NaCl.

The strain may also be characerized as a chemo-organotroph, since it grew on complex substrates such as yeast extract and peptone; and hydrolyzed starch, gelatin, casein, carboxymethylcellulose and amorphous cellulose.

The strain was observed to have metabolism that was respiratory and also fermentative. Acid was produced both aerobically and anaerobically from (API 50CH): L-arabinose, D-xylose, D-glucose, D-fructose, D-mannose, rhamnose (weak), cellobiose, maltose, sucrose, trehalose, gentiobiose, D-turanose, D-lyxose and 5-keto-gluconate (weak). Amygdalin, arbutin, salicin and esculin are also utilized. The strain was unable to utilize: ribose, lactose, galactose, melibiose, D-raffinose, glycogen, glycerol, erythritol, inositol, mannitol, sorbitol, xylitol, arabitol, gluconate and lactate.

The strain was observed to be susceptible to the antibiotics, ampicillin, chloramphenicol, erythromycin, fusidic acid, methicillin, novobiocin, streptomycin, tetracycline, sulphafurazole, oleandomycin, polymixin, rifampicin, vancomycin and bacitracin; but resistant to gentamicin, nitrofurantoin, nalidixic acid, sulphmethoxazole, trimethoprim, penicillin G, neomycin and kanamycin.

The following enzymes, aside from the protease of the present invention, were observed to be produced (ApiZym, API Coryne); C4-esterase, C8-esterase/lipase, leucine arylamidase, alpha-chymotypsin, alpha-glucosidase, beta-glucosidase and pyrazinamidase.

The strain was observed to exhibit the following chemotaxonomic characteristics. Major fatty acids (>10% of total) were C16:1 (28.1%), C18:0 (31.1%), C18:1 (13.9%). N-saturated (79.1%), n-unsaturated (19.9%). Fatty acids with even numbers of carbons accounted for 98%. Main polar lipid components: phosphatidylglycerol (PG) and 3 unidentified glycolipids (alpha-napthol positive); DPG, PGP, PI and PE were not detected. Menaquinones MK-4, MK-6, MK-7 and MK-9 were the main isoprenoids present. The cell wall peptidoglycan type was A4β with L-ornithine as diamino acid and D-aspartic acid in the interpeptide bridge.

With regard to toxicity evaluation, there are no known toxicity or pathogenicity issues associated with bacteria of the genus *Cellulomonas*.

Although there may be slight variations in the sequence of a naturally occurring enzyme within a given species of organism, enzymes of a specific type produced by organisms of the same species generally are substantially identical with respect to substrate specificity and/or proteolytic activity levels under various conditions (e.g., temperature, pH, water hardness, oxidative conditions, chelating conditions, and concentration), and the like. Thus for the purposes of the present invention, it is contemplated that other strains and species of *Cellulomonas* may also produce the *Cellulomonas* protease of the present invention and thus may be used as a source thereof.

The proteolytic polypeptides of this invention can be characterized either physicochemically or functionally, and/or by both. Physicochemical characterization takes advantages of well known techniques such as SDS electrophoresis, gel filtration, amino acid composition, mass spectroscopy (MALDI-TOF) and sedimentation to determine the molecular weight of a protein, isoelectric focusing to determine the pI of a protein, amino acid sequencing to determine the amino acid sequence of a protein, crystallography studies to determine the tertiary structure of a protein, and antibody binding to determine antigenic epitopes present in a protein.

Functional characteristics are determined by techniques well known to the practitioner in the protease field and include, but are not limited to, hydrolysis of various commercial substrates, e.g., di-methyl casein ("DMC") and/or AAPF-pNA. This preferred technique for functional characterization is described in greater detail in Examples 5 - 11, below.

In one aspect of the present invention, the protease has a molecular weight of about 17kD to about 21kD, for example about 18kD to 19kD, for example 18700 daltons to 18800 daltons, for example about 18776 daltons (as determined by MALDI-TOF). In another aspect of the present invention, the protease provides a MALDI-TOF spectrum as set forth in Fig. 17.

The mature protease also displays proteolytic activity, e.g., hydrolytic activity on a substrate having peptide linkages, e.g., DMC; and comparative or enhanced wash performance under identified conditions. In one aspect the protease displays at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% proteolytic activity as compared to the proteolytic activity of 69B4. In one aspect the protease displays at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% proteolytic activity as compared to the proteolytic activity of proteases sold under the tradenames SAVINASE or PURAFECT under the same conditions. In one aspect the protease displays comparative or enhanced wash performance under identified conditions as compared to 69B4 under the same conditions. In one aspect the protease displays comparative or enhanced wash performance under identified conditions as compared to proteases sold under the tradenames SAVINASE or PURAFECT under the same conditions.

According to an aspect of the invention, the protease and/or polynucleotide encoding the protease of the present invention is in a purified form, i.e., present in a particular composition in a higher or lower concentration than exists in a naturally occurring or wild type organism or in combination with components not normally present upon expression from a naturally occurring or wild type organism.

II. Obtaining polynucleotides encoding a *Cellulomonas* protease

The nucleic acid encoding a protease may be obtained by standard procedures known in the art from, for example, cloned DNA (*e.g.*, a DNA "library"), by chemical synthesis, by cDNA cloning, by PCR, or by the cloning of genomic DNA, or fragments thereof, purified from a desired cell, such as a bacterial or fungal species (*See*, for example, Sambrook *et al.*, 1989, Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; Glover, DM and Hames, BD (Eds.), DNA Cloning: A Practical Approach, *Vols 1 and 2*, Second Edition). Synthesis of polynucleotide sequences is well known in the art and can be found in, for example, Beaucage & Caruthers, Tetrahedron Letts. 22(20):1859-1862 (1981), e.g., using an automated synthesizesr, e.g., as described in Needham-VanDevanter, et al., Nucl Acids Res. 12:6159-

6168 (1984). DNA sequences can also be custom made and ordered from a variety of commercial sources, for example Promega (Madison, Wisc., USA). Nucleic acid sequences derived from genomic DNA may contain regulatory regions in addition to coding regions.

In the molecular cloning of the gene from genomic DNA, DNA fragments are generated, some of which will comprise at least a portion of the desired gene. The DNA may be cleaved at specific sites using various restriction enzymes. Alternatively, one may use DNAse in the presence of manganese to fragment the DNA, or the DNA can be physically sheared, as for example, by sonication. The linear DNA fragments can then be separated according to size by standard techniques, including but not limited to, agarose and polyacrylamide gel electrophoresis, PCR and column chromatography.

Once nucleic acid fragments are generated, identification of the specific DNA fragment encoding a protease may be accomplished in a number of ways. For example, a proteolytic hydrolyzing enzyme encoding the *asp* gene or its specific RNA, or a fragment thereof, such as a probe or primer, may be isolated and labeled and then used in hybridization assays to detect a generated gene. (Benton, W. and Davis, R., 1977, *Science 196*:180; Grunstein, M. and Hogness, D., 1975, *Proc. Natl. Acad. Sci. USA 72*:3961). Those DNA fragments sharing substantial sequence similarity to the probe will hybridize under medium to high stringency.

The process of amplification as carried out in polymerase chain reaction (PCR) technologies is described in Dieffenbach CW and GS Dveksler (1995, PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview NY). A nucleic acid sequence of at least about 4 nucleotides and as many as about 60 nucleotides from SEQ ID NO:1, 2, 3 or 4, preferably about 12 to 30 nucleotides, and more preferably about 25 nucleotides can be used as a probe or PCR primer.

A method of isolating a nucleic acid construct of the invention from a cDNA or genomic library is by use of polymerase chain reaction (PCR) using degenerate oligonucleotide probes prepared on the basis of the amino acid sequence of the protein having the amino acid sequence as shown in SEQ ID NOS:1 -5. The probes can be of any segment length, for example at least 4, at least 5, at least 8, at least 15, at least 20, nucleotides in length. Exemplary probes in the present application utilized a primer comprising a TTGWHCGT and a GDSGG polynucleotide sequence as more fully described in Example Section 2B, entitled "PCR amplification of a serine protease gene fragment". Other exemplary methodology includes the techniques described in US Patent No. 4,683,202.

In view of the above, it will be appreciated that the polynucleotide sequences can be provided and based on the polynucleotide sequence provided in Figs.1-5 are useful for obtaining identical or homologous fragments of polynucleotides from other species, and particularly from bacteria which encode enzymes having serine protease activity.

III. Expression and recovery of serine protease enzymes

Many methods for cloning a polypeptide, for example a *Cellulomonas* derived polypeptide having proteolytic activity or an additional enzyme (e.g., a second peptide having proteolytic activity, e.g., a protease, or a cellulase, a mannanase, an amylase, etc.), and for introducing, into the genes or genome, multiple copies of the polynucleotides that encode the polypeptides (e.g. *asp* , subtilase genes, *Cellulomonas* cellulase genes) are well known in the art.

In general, standard procedures for cloning of genes and introducing exogenous proteases encoding regions (including multiple copies of the exogenous encoding regions) into said genes may be used in order to obtain a *Cellulomonas* 69B4 protease derivative or homologue thereof. For further description of suitable techniques reference is made to working examples herein (vide infra) and (Sambrook et al. (1989) Molecular cloning: A laboratory manual, Cold Spring Harbor lab., Cold Spring Harbor, N.Y.; Ausubel, F. M. et al. (eds.) "Current protocols in Molecular Biology". John Wiley and Sons, 1995; Harwood, C. R., and Cutting, S. M. (eds.) "Molecular Biological Methods for Bacillus". John Wiley and. Sons, 1990); and WO 96/34946.

The polynucleotide sequences of the present invention may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employed in that expression vector to transform an appropriate host according to techniques well established in the art. The polypeptides produced on expression of the DNA sequences of this invention can be isolated from the fermentation of cell cultures and purified in a variety of ways according to well established techniques in the art. One of skill in the art is capable of selecting the most appropriate isolation and purification techniques.

More particularly, the present invention provides constructs, vectors comprising polynucleotides described herein, host cells transformed with such vectors, proteases expressed by such host cells, expression methods and systems for the production of serine protease enzymes derived from microorganisms, such as *Cellulomonas* species. Once a polynucleotide encoding a serine protease of the present invention is obtained, recombinant host cells containing the polynucleotide may be constructed using techniques well known in

the art. Molecular biology techniques are disclosed in Sambrook et al., Molecular Biology Cloning: A Laboratory Manual, Second Edition (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).

IV. Recombinant Vector

One aspect of the present invention encompasses a vector comprising the aforementioned polynucleotides. The polynucleotide construct of the invention encoding the protease may suitably be of genomic origin, for instance obtained by preparing a genomic library and screening for DNA sequences coding for all or part of the protease by hybridization using synthetic oligonucleotide probes in accordance with standard techniques (cf. Sambrook et al., Molecular Cloning, A Laboratory Mannual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1989). The polynucleotide construct is preferably a DNA construct which term will be used exclusively in the following. For the present purpose, the DNA sequence encoding the protease is obtainable by isolating chromosomal DNA from the *Cellulomonas* strain 69B4 described in Example I, supra and amplifying the sequence by PCR methodology.

The nucleic acid construct of the invention encoding the protease may also be prepared synthetically by established standard methods, e.g., the phosphoramidite method described by Beaucage and Caruthers, Tetrahedron Letters 22 (1981), 1859-1869, or the method described by Matthes et al., EMBO Journal 3 (1984), 801-805. According to the phosphoramidite method, oligonucleotides are synthesized, e.g., in an automatic DNA synthesizer, purified, annealed, ligated and cloned in suitable vectors..

Furthermore, the nucleic acid construct may be of mixed synthetic and genomic origin and may be prepared by ligating fragments of synthetic or genomic DNA (as appropriate), the fragments corresponding to various parts of the entire nucleic acid construct, in accordance with standard techniques.

In a further aspect, the present invention relates to a vector comprising a DNA construct of the invention. In another aspect, the present invention encompasses recombinant vectors. The recombinant vector into which the DNA construct of the invention is inserted may be any vector which may conveniently be subjected to recombinant DNA procedures, and the choice of vector will often depend on the host cell into which it is to be introduced. Thus, the vector may be an autonomously replicating vector, i.e., a vector which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g., a plasmid. Alternatively, the vector may be one which, when

introduced into a host cell, is integrated into the host cell genome and replicated together with the chromosome(s) into which it has been integrated.

A variety of vectors and transformation and expression cassettes suitable for the cloning, transformation and expression in fungus, yeast, bacteria, insect and plant cells are known by those of skill in the art. Typically, the vector or cassette contains sequences directing transcription and translation of the nucleic acid, a selectable marker, and sequences allowing autonomous replication or chromosomal integration. Suitable vectors comprise a region 5' of the gene which harbors transcriptional initiation controls and a region 3' of the DNA fragment which controls transcriptional termination. These control regions may be derived from genes homologous or heterologous to the host as long as the control region selected is able to function in the host cell.

The vector is preferably an expression vector in which the DNA sequence encoding the protease of the invention is operably linked to additional segments required for transcription of the DNA. In general, the expression vector is derived from plasmid or viral DNA, or may contain elements of both. Exemplary vectors include pSEGCT, pSEACT and / or pSEA4CT. Construction of such vectors is well known in the art , e.g., US Patent No. 6,287,839 and/or International Publication WO 02/50245. In one aspect the vector pSEGCT (about 8302 bp) is useful in the construction of a vector comprising the polynucleotides described herein, e.g., pSEG69B4T (Fig. 18)

The additional segments required for transcription may include regulatory segments, e.g., promoters, secretory segments, inhibitors, global regulators, etc., as known in the art. An exemplary promoter may be any DNA sequence which shows transcriptional activity in the host cell of choice and may be derived from genes encoding proteins either homologous or heterologous to the host cell. Specifically, examples of suitable promoters for use in bacterial host cells include but are not limited to the promoter of the *Bacillus stearothermophilus* maltogenic amylase gene, the *Bacillus amyloliquefaciens* (BAN) amylase gene, the *Bacillus subtilis* alkaline protease gene, the *Bacillus pumilus* xylosidase gene, the *Bacillus thuringiensis* cryIIIA or the *Bacillus licheniformis* alpha-amylase gene. Other promoters that may be used include but are not limited to phage Lambda P_R or P_L promoters or the *E. coli* lac, trp or tac promoters.

The promoter may also be derived from a gene encoding said protease or a fragment thereof having substantially the same promoter activity as said sequence. The invention further encompasses nucleic acid sequences which hybridize to the promoter sequence under the intermediate, high or maximum stringency conditions, or which have at

least about 90% homology and preferably about 95% homology to such promoter, but which have substantially the same promoter activity as said sequence. This promoter may be used to promote the expression of either said protease or a heterologous DNA sequence, e.g. other enzymes as described later in this application.

The recombinant vector of the invention may further comprise a DNA sequence enabling the vector to replicate in the host cell in question. When the host cell is a bacterial cell, sequences enabling the vector to replicate are various ori sequences.

The vector may also comprise a selectable marker, e.g., a gene the product of which which confers resistance to a drug, e.g., ampicillin, kanamycin, tetracycline, chloramphenicol, neomycin, hygromycin or methotrexate.

To direct a polypeptide of the present invention into the secretory pathway of the host cells, a secretory signal sequence (also known as a leader sequence or pre sequence) may be provided in the recombinant vector. The secretory signal sequence is joined to the DNA sequence encoding the precursor protease in the correct reading frame (see, for example, FIGS. 1A-B and 2A-B). Depending on whether the protease is to be expressed intracellularly or is secreted, a polynucleotide sequence or expression vector of the invention can be engineered with or without a natural polypeptide signal sequence or a signal sequence which functions in bacteria (e.g., *Bacillus subtilis*), fungus (e.g., *Aspergillus niger*), other prokaryoktes or eukaryotes. Expression ca also be achieved by either removing or partially removing said signal sequence

For secretion from bacterial cells, the signal peptide may be a naturally occurring signal peptide, or a functional part thereof, or it may be a synthetic peptide. Suitable signal peptides include but are not limited to sequences derived from *Bacillus licheniformis* alpha-amylase, *Bacillus lentus* alkaline protease, and *Bacillus amyloliquefaciens* amylase. Another suitable signal sequence is the signal peptide derived from *Cellulomonas* strain 69B4, as described supra.

For secretion from bacterial cells, the signal peptide may also be the signal peptide from the protease disclosed in the instant application. This signal may be used to facilitate the secretion of either said protease or a heterologous DNA sequence, e.g. a second protease, e.g. wild-type protease, a BPN' variant protease, a GG36 variant protease, a lipase, a cellulase, a mananase, etc. These second enzymes may be encoded by the DNA sequence and/or the amino acid sequences described in the art, e.g., US Patents 6,465,235 (WO 98/22500); 6,287,839 (WO 92/05249); 5,965,384 (EP 0305216B1); 5,795,764 (WO 94/25576). Furthermore, it is contemplated that the signal sequence peptide may also be

operatively linked to an endogenous sequence to activate and secrete such endogenous encoded protease.

The procedures used to ligate the DNA sequences coding for the present protease, the promoter and/or secretory signal sequence, respectively, and to insert them into suitable vectors containing the information necessary for replication, are well known to persons skilled in the art (cf., for instance, Sambrook et al., op. cid.).

The nucleic acid construct may also be prepared by polymerase chain reaction using specific primers, for instance as described in U.S. Pat. No. 4,683,202 or Saiki et al., Science 239 (1988), 487-491.

V. Host cells

In another aspect of the present invention encompasses host cell transformed with the vectors described above. The polynucleotide encoding the protease of the present invention that is introduced into the host cell may be either homologous or heterologous to the host in question. If homologous to the host cell, e.g., additional copies of the native protease produced by the host cell, it may be operably connected to another homologous or heterologous promoter sequence or, if applicable, another secretory signal sequence and/or terminator sequence than in its natural environment. Thus, the polypeptide DNA sequence may be multiple copies of a homologous polypeptide sequence, a heterologous polypeptide sequence from another organism, or it may be a synthetic polypeptide sequence.

The host cell into which the DNA construct or the recombinant vector of the invention is introduced may be any cell which is capable of producing the present alkaline protease and includes but is not limited to bacteria, yeast, fungi and higher eukaryotic cells.

Examples of bacterial host cells which, on cultivation, are capable of producing and secreting the protease of the invention are Gram positive bacteria such as strains of *Bacillus*, *Streptomyces*, or *Thermobifida*, for example strains of *B. subtilis*, *B. licheniformis*, *B. lentus*, *B. brevis*, *B. stearothermophilus*, *B. alkalophilus*, *B. amyloliquefaciens*, *B. coagulans*, *B. circulans*, *B. lautus*, *B. megatherium*, *B. thuringiensis*, *S. griseus*; *S. lividans*, *S. coelicolor*; *S. avermitilis* and *T. fusca* or Gram negative bacteria such as *Escherichia coli*. In another aspect the host cell is selected from *S. lividans* TK23 and/or TK21. Examples of fungal host cells which, on cultivation, are capable of producing and secreting the protease of the invention are the fungi of *Trichoderma* spp. and *Aspergillus* spp. In another aspect, the *Trichoderma* spp. is *Trichoderma resei*. In another aspect , the *Aspergillus* spp. is *Aspergillus niger*. The transformation of the bacteria may be effected by protoplast

transformation or by using competent cells in a manner known per se (cf. Sambrook et al., supra). A preferred general transformation and expression protocol for protease deleted strains is provided in Ferrari, et al., U.S. Patent No. 5,264,366, incorporated by reference. Transformation and expression in *Aspergillus* is described in, for example, Berka, et al., U.S. Patent 5,364,770, incorporated by reference. Transformation and expression in *Streptomyces* can be found in Hopwood, et al., GENETIC MAIPULATION OF STREPTOMYCES: LABORATORY MANUAL, Innis (1985), which is incorporated by reference. Transformation and expression in *Streptomyces lividans* is described in Fernandez-Abalos, J., et al, Microbiology 149:1623-1632 (2003) and in Example 4 herein.

The transformation of the polynucleotide sequences of this invention into *Bacillus* may be ineffective in terms of expression. Thus, in one aspect, when transforming *Bacillus* spp., to utilize the aprE promoter in combination with known *Bacillus*-derived signal and other regulatory sequences. When the transformation host cell is *Aspergillus*, one aspect of the invention utilizes the glaA promoter. With *Streptomyces*, in one aspect of the present invention, the promoter utilized is the Glucose Isomerase (GI) promoter of *Actinoplanes missouriensis*.

When expressing the protease in bacteria such as *E. coli*, the protease may be retained in the cytoplasm, typically as insoluble granules (known as inclusion bodies), or may be directed to the periplasmic space by a bacterial secretion sequence. In the former case, the cells are lysed and the granules are recovered and denatured after which the protease is refolded by diluting the denaturing agent. In the later case, the protease may be recovered from the periplasmic space by disrupting the cells, e.g., by sonication or osmotic shock, to release the contents of the periplasmic space and recovering the protease.

The transformed host cell described above is then cultured in a suitable nutrient medium under conditions permitting the expression of the present protease, after which the resulting protease is recovered from the culture. The medium used to culture the cells may be any conventional medium suitable for growing the host cells, such as minimal or complex media containing appropriate supplements. Suitable media are available from commercial suppliers or may be prepared according to published recipes (e.g., in catalogues of the American Type Culture Collection). The protease produced by the cells may then be recovered from the culture medium by conventional procedures including separating the host cells from the medium by centrifugation or filtration, precipitating the proteinaceous components of the supernatant or filtrate by means of a salt, e.g., ammonium sulphate, purification by a variety of chromatographic procedures, e.g., ion exchange

chromatography, gel filtration chromatography, affinity chromatography, or the like, dependent on the type of protease in question.

VI Applications of serine protease enzymes

Applicants have discovered that novel proteases according to the present invention have enhanced thermal stability, enhanced oxidative stability, and enhanced chelator stability and therefore useful in cleaning compositions. In addition, under certain wash conditions, such protease exhibits comparative or enhanced wash performance as compared with currently used subtilisin proteases. The cleaning and/or enzyme compositions of the present invention may take on a variety of cleaning compositions. The identified proteases can also be utilized in the same manner that subtilisin proteases have heretofore been used, in animal feed applications, leather processing, e.g., bating, protein hydrolysis and in textile uses. The identified proteases can also be utilized in the same manner that proteases have heretofore been used, in personal care applications. An aspect of this invention, the protease, the polypeptide expressing or encoding the same, vectors, host cells and uses of such protease are described below.

It is contemplated that the protease of the invention may be used in a number of industrial applications, in particular within the cleaning, disinfecting, animal feed or textile industries. The protease of this invention may be combined with detergents, builders, bleaching agents and other conventional ingredients to produce a variety of novel cleaning compositions useful in the laundry and other cleaning arts such as, for example, laundry detergents (both powdered and liquid), laundry pre-soaks, all fabric bleaches, automatic dishwashing detergents (both liquid and powdered), household cleaners, particularly bar and liquid soap applications, and drain openers. In addition, the protease may also be employed in the cleaning of contact lenses and other membranes by contacting such materials with an aqueous solution of the cleaning composition. In addition these naturally occurring proteases can be used, for example in peptide hydrolysis, waste treatment, textile applications, medical device cleaning, biofilm removal and as fusion-cleavage enzymes in protein production, etc. The composition of such products is not critical to this invention, and the same may be readily prepared by combining a cleaning effective amount of the protease or an enzyme composition comprising the protease enzyme preparation with the conventional components of such compositions in their art recognized amounts.

The considerations arising out of detergent applications provide insight into the beneficial use of proteases in these applications.

For example, there are a variety of wash conditions including varying detergent formulations, wash water volume, wash water temperature and length of wash time that a protease variant might be exposed to. For example, detergent formulations used in different areas have different concentrations of their relevant components present in the wash water. For example, a European detergent typically has about 4500-5000 ppm of detergent components in the wash water while a Japanese detergent typically has approximately 667 ppm of detergent components in the wash water. In North America, particularly the United States, a detergent typically has about 975 ppm of detergent components present in the wash water.

A low detergent concentration system includes detergents where less than about 800 ppm of detergent components are present in the wash water. Japanese detergents are typically considered low detergent concentration system as they have approximately 667 ppm of detergent components present in the wash water.

A medium detergent concentration includes detergents where between about 800 ppm and about 2000ppm of detergent components are present in the wash water. North American detergents are generally considered to be medium detergent concentration systems as they have approximately 975 ppm of detergent components present in the wash water. Brazil typically has approximately 1500 ppm of detergent components present in the wash water.

A high detergent concentration system includes detergents where greater than about 2000 ppm of detergent components are present in the wash water. European detergents are generally considered to be high detergent concentration systems as they have approximately 4500-5000 ppm of detergent components in the wash water.

Latin American detergents are generally high suds phosphate builder detergents and the range of detergents used in Latin America can fall in both the medium and high detergent concentrations as they range from 1500 ppm to 6000 ppm of detergent components in the wash water. As mentioned above, Brazil typically has approximately 1500 ppm of detergent components present in the wash water. However, other high suds phosphate builder detergent geographies, not limited to other Latin American countries, may have high detergent concentration systems up to about 6000 ppm of detergent components present in the wash water.

In light of the foregoing, it is evident that concentrations of detergent compositions in typical wash solutions throughout the world varies from less than about 800 ppm of detergent composition ("low detergent concentration geographies"), for example about 667

ppm in Japan, to between about 800 ppm to about 2000 ppm ("medium detergent concentration geographies"), for example about 975 ppm in U.S. and about 1500 ppm in Brazil, to greater than about 2000 ppm ("high detergent concentration geographies"), for example about 4500 ppm to about 5000 ppm in Europe and about 6000 ppm in high suds phosphate builder geographies.

The concentrations of the typical wash solutions are determined empirically. For example, in the U.S., a typical washing machine holds a volume of about 64.4 L of wash solution. Accordingly, in order to obtain a concentration of about 975 ppm of detergent within the wash solution about 62.79 g of detergent composition must be added to the 64.4 L of wash solution. This amount is the typical amount measured into the wash water by the consumer using the measuring cup provided with the detergent.

As a further example, different geographies use different wash temperatures. The temperature of the wash water in Japan is typically less than that used in Europe. For example, the temperature of the wash water in North America and Japan can be between 10 and 30 degrees centigrade, for example about 20 degrees C, whereas the temperature of wash water in Europe is typically between 30 and 60 degrees centigrade, for example about 40 degrees C.

As a further example, different geographies may have different water hardness. Water hardness is typically described as grains per gallon mixed Ca^{2+}/Mg^{2+}. Hardness is a measure of the amount of calcium (Ca^{2+}) and magnesium (Mg^{2+}) in the water. Most water in the United States is hard, but the degree of hardness varies. Moderately hard (60-120 ppm) to hard (121-181 ppm) water has 60 to 181 parts per million [parts per million converted to grains per U.S. gallon is ppm # divided by 17.1 equals grains per gallon] of hardness minerals.

Water	Grains per gallon	Parts per million
Soft	less than 1.0	less than 17
Slightly hard	1.0 to 3.5	17 to 60
Moderately hard	3.5 to 7.0	60 to 120
Hard	7.0 to 10.5	120 to 180
Very hard	greater than 10.5	greater than 180

European water hardness is typically greater than 10.5 (for example 10.5-20.0) grains per

gallon mixed Ca^{2+}/Mg^{2+}, for example about 15 grains per gallon mixed Ca^{2+}/Mg^{2+}. North American water hardness is typically greater than Japanese water hardness, but less than European water hardness. For example, North American water hardness can be between 3 to10 grains, 3-8 grains or about 6 grains. Japanese water hardness is typically the lower than North American water hardness, typically less than 4, for example 3 grains per gallon mixed Ca^{2+}/Mg^{2+}.

Accordingly one aspect of the present invention includes a protease that shows surprising wash performance in at least one set of wash conditions as related to water temperature, water hardness and/or detergent concentration. In another aspect of the present invention, the protease of the present invention exhibited comparative wash performance as compared to subtilisin proteases. In another aspect of the present invention the protease of the present invention exhibited enhanced wash performance as compared to subtilisin proteases. Accordingly one aspect of the present invention includes a protease that shows enhanced oxidative stability, enhanced thermal stability, and/or enhanced chelator stability.

Accordingly another aspect of the invention includes a *Cellulomonas* protease isolated according to the present invention as being useful in applications in which it is desired to clean protein based stains from textiles or fabrics.

The "compact" form of the cleaning compositions herein is best reflected by density and, in terms of composition, by the amount of inorganic filler salt; inorganic filler salts are conventional ingredients of detergent compositions in powder form; in conventional detergent compositions, the filler salts are present in substantial amounts, typically 17-35% by weight of the total composition. In the compact compositions, the filler salt is present in amounts not exceeding 15% of the total composition, in another aspect not exceeding 10%, in another aspect not exceeding 5% by weight of the composition. The inorganic filler salts, in the context of the present invention can selected from the alkali and alkaline-earth-metal salts of sulfates and chlorides. A preferred filler salt is sodium sulfate.

The cleaning compositions of the invention may for example, be formulated as hand and machine laundry detergent compositions including laundry additive compositions and compositions suitable for use in the pretreatment of stained fabrics, rinse added fabric softener compositions, and compositions for use in general household hard surface cleaning operations and dishwashing operations. The art is familiar with the different formulations which can be used as cleaning compositions. The proteases of the present invention may comprise comparative or enhanced performance in a detergent composition (as compared

to other protease, e.g., naturally occurring or variant subtilisin proteases). Cleaning performance can be determined by comparing the proteases of the present invention with those subtilisin proteases in various cleaning assays concerning enzyme sensitive stains such as egg, grass, blood or milk as determined by usual spectrophotometric or analytical methodologies after standard wash cycle conditions. Exemplary assays are described in WO 99/34011, US Patent No. 6,605,458, Example 3 [detergent dose of 3.0 g/l at pH10.5, wash time 15 minutes, at 15 C, water hardness of 6°dH, 10nM enzyme concentration in 150 ml glass beakers with stirring rod, 5 textile pieces (phi 2.5 cm) in 50 ml, EMPA 117 test material from Center for Test Materials Holland. Measurement of reflectance "R" on the test material done at 460 nm using a Macbeth ColorEye 7000 photometer], See also Experiment 5, eentitled "Determination of Cleaning Activity, in this application).

The addition of proteases of the invention to conventional cleaning compositions does not create any special use limitation. In other words, any temperature and pH suitable for the detergent is also suitable for the present compositions as long as the pH is within the above range, and the temperature is below the described protease's denaturing temperature. In addition, proteases of the invention can be used in a cleaning composition without detergents, again either alone or in combination with builders and stabilizers.

When used in cleaning compositions or detergents, oxidative stability is a further consideration which characterizes the naturally occurring proteases described in the examples. The stability may be enhanced, diminished, or comparable to subtilisin proteases as is desired for various uses.

When used in cleaning compositions or detergents, thermal stability is a further consideration which characterizes the naturally occurring proteases described in the examples. The stability may be enhanced, diminished, or comparable to subtilisin proteases as is desired for various uses.

When used in cleaning compositions or detergents, chelator stability is a further consideration which characterizes the naturally occurring proteases described in the examples. The stability may be enhanced, diminished, or comparable to subtilisin proteases as is desired for various uses.

In an aspect of the present invention, naturally occurring proteases are provided which exhibit modified enzymatic activity at different pH's when compared to subtilisin proteases. A pH-activity profile is a plot of pH against enzyme activity and may be constructed as illustrated in Example 10 and 11 or by methods known in the art. It may be desired to obtain naturally occurring proteases with broader profiles, i.e., those having

greater activity at certain pH than a comparable subtilisin protease, but no significantly greater activity at any pH, or naturally occurring homologues with sharper profiles, i.e. those having enhanced activity when compared to subtilisin proteases at a given pH, and lesser activity elsewhere.

In one aspect of the present invention, the cleaning compositions of the present invention can comprise, proteases of the present invention at a level from 0.00001 % to 10% of *Cellulomonas* protease by weight of the composition and the balance, e.g., 99.999% to 90.0% of cleaning adjunct materials by weight of composition. In other aspects of the present invention, the cleaning compositions of the present invention also comprise, proteases at a level of 0.0001 % to 10%, 0.001% to 5%, 0.001% to 2%, 0.005% to 0.5% *Cellulomonas* protease by weight of the composition and the balance of the cleaning composition, e.g., 99.9999% to 90.0%, 99.999 % to 98%, 99.995% to 99.5% by weight of the cleaning composition comprise cleaning adjunct materials.

Preferred cleaning compositions, in addition to the protease preparation of the invention, may comprise one or more additional enzymes or enzyme derivatives which provide cleaning performance and/or fabric care benefits. Such enzymes include other proteases, lipases, cutinases, amylases, cellulases, peroxidases, oxidases (e.g. laccases); and/or mannanases.

Other Proteases: Any other protease suitable for use in alkaline solutions can be used. Suitable proteases include those of animal, vegetable or microbial origin. Microbial origin is preferred. Chemically or genetically modified mutants are included. The protease may be a serine protease, preferably an alkaline microbial protease or a trypsin-like protease. Examples of alkaline proteases are subtilisins, especially those derived from *Bacillus*, e.g., subtilisin, *lentus*, *amyloliquefaciens*, subtilisin Carlsberg, subtilisin 309, subtilisin 147 and subtilisin 168. Additional examples include those mutant proteases described in US Patents RE 34,606; 5,955,340; 5,700,676; 6,312,936 and 6,482,628. Examples of trypsin like proteases are trypsin (e.g. of porcine or bovine origin) and the *Fusarium* protease described in WO 89/06270.

Preferred commercially available protease enzymes include those sold under the trade names MAXATASE, MAXACAL, MAXAPEM; OPTICLEAN, OPTIMASE, PROPERASE, PURAFECT and PURAFECT OXP by Genencor International, those sold under the trade names ALCALASE, SAVINASE, PRIMASE, DURAZYM, RELASE and ESPERASE by Novo Nordisk A/S (Denmark); and those sold under the trade names BLAP by Henkel Kommanditgesellschaft auf Aktien, Duesseldorf, Germany (WO95/23221 [US

Patent Nos. 5,801,039, 5,855,625]; WO 92/21760 (US Patent Nos. 5,340,735, 5,500,364)). An additional BPN' variant (BPN'-var 1) is described in US RE 34,606. An additional GG36-variant (GG36-var.1) is described in US 5,955,340 and 5,700,676. An additional GG36-variant is described in US Patents 6,312,936 and 6,482,628. In one aspect of the present invention, the cleaning compositions of the present invention can comprise, additional protease enzymes at a level from 0.00001 % to 10% of additional protease by weight of the composition and 99.999% to 90.0% of cleaning adjunct materials by weight of composition. In other aspects of the present invention, the cleaning compositions of the present invention also comprise, proteases at a level of 0.0001 % to 10%, 0.001% to 5%, 0.001% to 2%, 0.005% to 0.5% *Cellulomonas* protease by weight of the composition and the balance of the cleaning composition, e.g., 99.9999% to 90.0%, 99.999 % to 98%, 99.995% to 99.5% by weight of the cleaning compositon comprise cleaning adjunct materials.

Lipases: Any lipase suitable for use in alkaline solutions can be used. Suitable lipases include those of bacterial or fungal origin. Chemically or genetically modified mutants are included. Examples of useful lipases include a *Humicola lanuginosa* lipase, e.g., as described in EP 258 068 and EP 305 216, a *Rhizomucor miehei* lipase, e.g., as described in EP 238 023, a *Candida lipase*, such as a *C. antarctica* lipase, e.g., the *C. antarctica* lipase A or B described in EP 214 761, a *Pseudomonas* lipase such as a *P. alcaligenes* and *P. pseudoalcaligenes* lipase, e.g., as described in EP 218 272, a *P. cepacia* lipase, e.g., as described in EP 331 376, a *P. stutzeri* lipase, e.g., as disclosed in GB 1,372,034, a *P. fluorescens* lipase, a *Bacillus lipase*, e.g., a *B. subtilis* lipase (Dartois et al., (1993), Biochemica et Biophysica acta 1131, 253-260), a *B. stearo-thermophilus* lipase (JP 64/744992) and a *B. pumilus* lipase (WO 91/16422).

Furthermore, a number of cloned lipases may be useful, including the *Penicillium camembertii* lipase described by Yamaguchi et al., (1991), Gene 103, 61-67), the *Geotricum candidum* lipase (Schimada, Y. et al., (1989), J. Biochem., 106, 383-388), and various *Rhizopus* lipases such as a *R. delemar* lipase (Hass, M. J et al., (1991), Gene 109, 117-113), a *R. niveus* lipase (Kugimiya et al., (1992), Biosci. Biotech. Biochem. 56, 716-719) and a *R. oryzae* lipase.

Other types of lipolytic enzymes such as cutinases may also be useful, e.g., a cutinase derived from *Pseudomonas mendocina* as described in WO 88/09367, or a cutinase derived from *Fusarium solani pisi* (e.g. described in WO 90/09446).

Suitable lipases are lipases such as M1 LIPASE.TM., LUMA FAST TM. and LIPOMAX TM. (Genencor), LIPOLASE.TM. and LIPOLASE ULTRA TM. (Novo Nordisk A/S,

Denmark), and LIPASE P "Amano" (Amano Pharmaceutical Co. Ltd., Japan).

In one aspect of the present invention, the cleaning compositions of the present invention can further comprise lipases at a level from 0.00001 % to 10% of additional protease by weight of the composition and the balance of cleaning adjunct materials by weight of composition. In other aspects of the present invention, the cleaning compositions of the present invention also comprise, lipases at a level of 0.0001 % to 10%, 0.001% to 5%, 0.001% to 2%, 0.005% to 0.5% lipase by weight of the composition.

Amylases: Any amylase (.alpha. and/or .beta.) suitable for use in alkaline solutions can be used. Suitable amylases include those of bacterial or fungal origin. Chemically or genetically modified mutants are included. Amylases include, for example, a-amylases obtained from a special strain of *B. licheniformis*, described in more detail in GB 1,296,839. Commercially available amylases are DURAMYL TM., TERMAMYL.TM., FUNGAMYL.TM. and BAN.TM. (available from Novo Nordisk A/S) and RAPIDASE TM. and MAXAMYL P.TM. (available from Genencor International).

In one aspect of the present invention, the cleaning compositions of the present invention can further comprise amylases at a level from 0.00001 % to 10% of additional protease by weight of the composition and the balance of cleaning adjunct materials by weight of composition. In other aspects of the present invention, the cleaning compositions of the present invention also comprise, amylases at a level of 0.0001 % to 10%, 0.001% to 5%, 0.001% to 2%, 0.005% to 0.5% lipase by weight of the composition.

Cellulases: Any cellulase suitable for use in alkaline solutions can be used. Suitable cellulases include those of bacterial or fungal origin. Chemically or genetically modified mutants are included. Suitable cellulases are disclosed in U.S. Pat. No. 4,435,307, which discloses fungal cellulases produced from *Humicola insolens*. Especially suitable cellulases are the cellulases having color care benefits. Examples of such cellulases are cellulases described in European patent application No. 0 495 257.

Commercially available cellulases include CELLUZYME.TM. produced by a strain of *Humicola insolens*. (Novo Nordisk A/S), and KAC-500(B).TM. (Kao Corporation). Cellulases can also be incorporated as portions or fragments of mature wild-type or variant cellulases, wherein a portion of the N-terminus is deleted as described in US Patent No. 5,874,276.

In one aspect of the present invention, the cleaning compositions of the present invention can further comprise cellulases at a level from 0.00001 % to 10% of additional protease by weight of the composition and the balance of cleaning adjunct materials by weight of composition. In other aspects of the present invention, the cleaning compositions

of the present invention also comprise cellulases at a level of 0.0001 % to 10%, 0.001% to 5%, 0.001% to 2%, 0.005% to 0.5% lipase by weight of the composition.

Mananases . Any mananase suitable for use in detergent compositions and or alkaline solutions can be used. Suitable mananases include those of bacterial or fungal origin. Chemically or genetically modified mutants are included. Suitable mananases are disclosed in U.S. Pat. No. 6,566,114, which discloses *Bacillus* mananases produced from a cloned *E. coli* host cell; US. Pat. No.6,602,842, which discloses a *Bacillus* mannanase used in detergent compositions; US Patent No. 6,440,991 which discloses a cleaning composition comprising a mannanase enzyme.

In one aspect of the present invention, the cleaning compositions of the present invention can further comprise mananases at a level from 0.00001 % to 10% of additional protease by weight of the composition and the balance of cleaning adjunct materials by weight of composition. In other aspects of the present invention, the cleaning compositions of the present invention also comprise, mananases at a level of 0.0001 % to 10%, 0.001% to 5%, 0.001% to 2%, 0.005% to 0.5% mananase by weight of the composition.

Peroxidases/Oxidases: Peroxidase enzymes are used in combination with hydrogen peroxide or a source thereof (e.g. a percarbonate, perborate or persulfate). Oxidase enzymes are used in combination with oxygen. Both types of enzymes are used for "solution bleaching", i.e. to prevent transfer of a textile dye from a dyed fabric to another fabric when said fabrics are washed together in a wash liquor, preferably together with an enhancing agent as described in e.g. WO 94/12621 and WO 95/01426. Suitable peroxidases/oxidases include those of plant, bacterial or fungal origin. Chemically or genetically modified mutants are included.

In one aspect of the present invention, the cleaning compositions of the present invention can further comprise peroxidase and/or oxidase enzymes at a level from 0.00001 % to 10% of additional protease by weight of the composition and the balance of cleaning adjunct materials by weight of composition. In other aspects of the present invention, the cleaning compositions of the present invention also comprise, peroxidase and/or oxidase enzymes at a level of 0.0001 % to 10%, 0.001% to 5%, 0.001% to 2%, 0.005% to 0.5% peroxidase and/or oxidase enzymes by weight of the composition.

Mixtures of the above mentioned enzymes are encompassed herein, in particular a mixture of a *Cellulomonas* protease, a second protease, an amylase, a lipase, a mannanase and/or a cellulase.

It is contemplated by the inventors that the varying levels of the protease and one or

more additional enzymes may both independently range to 10%, the balance of the cleaning composition being cleaning adjunct materials. The specific selection of cleaning adjunct materials are readily made by considering the surface, item or fabric to be cleaned, and the desired form of the composition for the cleaning conditions during use (e.g., through the wash detergent use).

Examples of suitable cleaning adjunct materials include, but are not limited to, surfactants, builders, bleaches, bleach activators, bleach catalysts, other enzymes, enzyme stabilizing systems, chelants, optical brighteners, soil release polymers, dye transfer agents, dispersants, suds suppressors, dyes, perfumes, colorants, filler salts, hydrotropes, photoactivators, fluorescers, fabric conditioners, hydrolyzable surfactants, perservatives, anti-oxidants, anti-shrinkage agents, anti-wrinkle agents, germicides, fungicides, color speckles, silvercare, anti-tarnish and/or anti-corrosion agents, alkalinity sources, solubilizing agents, carriers, processing aids, pigments and pH control agents as described in U.S. Pat. Nos. 6,610,642, 6,605,458, 5,705,464, 5,710,115, 5,698,504, 5,695,679, 5,686,014 and 5,646,101. Specific cleaning composition materials are exemplified in detail hereinafter.

If the cleaning adjunct materials are not compatible with the *Cellulomonas* protease or the *Cellulomonas* protease homologue(s) in the cleaning compositions, then suitable methods of keeping the cleaning adjunct materials and the protease separate (not in contact with each other) until combination of the two components is appropriate can be used. Suitable methods can be any method known in the art, such as gelcaps, encapulation, tablets, physical separation, etc.

Preferably an effective amount of one or more *Cellulomonas* proteases described above are included in compositions useful for cleaning a variety of surfaces in need of proteinaceous stain removal. Such cleaning compositions include detergent compositions for cleaning hard surfaces, unlimited in form (e.g., liquid and granular); detergent compositions for cleaning fabrics, unlimited in form (e.g., granular, liquid and bar formulations); dishwashing compositions (unlimited in form and including both granular and liquid automatic dishwashing); oral cleaning compositions, unlimited in form (e.g., dentifrice, toothpaste and mouthwash formulations); and denture cleaning compositions, unlimited in form (e.g., liquid, tablet).

Several examples of various cleaning compositions wherein the protease of the present invention may be employed are discussed in further detail below.

The cleaning compositions of the present invention may be in the form of "fabric cleaning compositions" or "non-fabric cleaning compositions."

As used herein, "fabric cleaning compositions" include hand and machine laundry detergent compositions including laundry additive compositions and compositions suitable for use in the soaking and/or pretreatment of stained fabrics.

As used herein, "non-fabric cleaning compositions" include hard surface cleaning compositions, dishwashing detergent compositions, oral cleaning compositions, denture cleaning compositions and personal cleansing compositions.

When the cleaning compositions of the present invention are formulated as compositions suitable for use in a laundry machine washing method, the compositions of the present invention preferably contain both a surfactant and a builder compound and additionally one or more cleaning adjunct materials preferably selected from organic polymeric compounds, bleaching agents, additional enzymes, suds suppressors, dispersants, lime-soap dispersants, soil suspension and anti-redeposition agents and corrosion inhibitors. Laundry compositions can also contain softening agents, as additional cleaning adjunct materials.

The compositions of the present invention can also be used as detergent additive products in solid or liquid form. Such additive products are intended to supplement or boost the performance of conventional detergent compositions and can be added at any stage of the cleaning process.

When formulated as compositions for use in manual dishwashing methods the compositions of the invention preferably contain a surfactant and preferably other cleaning adjunct materials selected from organic polymeric compounds, suds enhancing agents, group II metal ions, solvents, hydrotropes and additional enzymes.

If needed the density of the laundry detergent compositions herein ranges from 400 to 1200 g/liter, preferably 500 to 950 g/liter of composition measured at 20.degree. C.

In view of the above discussion, a suitable detergent granular fabric cleaning composition in accordance with the invention may be prepared as described in Example 1, US Patent 6,605,458; a compact granular fabric cleaning composition may be prepared as described in Example 2, US Patent 6,605,458; a granular fabric cleaning composition useful in the laundering of colored fabrics may be prepared as described in Detergent Example III, US Patent 6,605,458; a granular fabric cleaning composition which provide softening through the wash capacity may be prepared as described in Detergent Example IV, US Patent 6,605,458; a Heavy duty liquid fabric cleaning composition in accordance with the invention may be prepared as described in Detergent Example 5, US Patent 6,605,458; A suitable hard surface cleaning compositon is that described in Examples 1-7 of US Patent

6,610,642, Examples 1-7 (Liquid Hard Surface Cleaning Compositions) of US Patent 6,376,450, and Examples 8-13 (Spray Compositions for Cleaning Hard Surfaces and Removing Household Mildew) of US Patent 6,376,450; a suitable dishwashing compostition is that described in Examples 14-25 of US Patent 6,610,642, Examples 14-24 (Dishwashing compositions) of US Patent 6,376,450, a suitable fabric cleaning compositon is that described in Examples 27-36 of US Patent 6,610,642 and Examples 27-47 of US Patent 6,376,450; a suitable granular laundry detergent composition of particular utility under European machine conditions is that described in Example 37 of US Patent 6,610,642; a suitable granular laundry detergent compositon of particular utility under Japanese conditions is described in Example 39 of US Patent 6,610,642; a suitable liquid fabric cleaning compositon is that described in Examples 40-45 of US Patent 6,610,642; suitable oral cleaning compositions are those described in Examples 48 and 49 (Dentifrice compositions) of US Patent 6,376,450; a suitable mouthwash compositon is described in Example 49 of US Patent 6,376,450;a suitable lozenge composition is described in Example 50 of US Patent 6,376,450; a suitable chewing gum composition is described in Example 51 of US Patent 6,376,450; and a suitable two-layer effervescent Denture Cleaning tablet formulation is described in US Patent 6,376,450. The formulations and descriptions of the compounds and cleaning adjunct materials contained in the aforementioned US Patent Nos. 6,376,450; 6,605,458; 6,605,458; and 6,610,642 are expressly incorporated by reference herein. Still further examples are set forth in Examples 12-17 in this application.

Still further, the present invention contemplates a method for the production of a food or animal feed, characterized in that protease according to the invention is mixed with said food or animal feed. Said protease can be added as a dry product before processing or as a liquid before or after processing. According to one aspect, wherein a dry powder is used, the enzyme is diluted as a liquid onto a dry carrier such as milled grain. The proteases of the present invention can be included in animal feed such as part of animal feed additives as described in, for example, US 5,612,055; US 5,314,692; and US 5,147,642.

The enzyme feed additive according to the present invention can be prepared in a number of ways. For instance, it can be prepared simply by mixing different enzymes having the appropriate activities to produce an enzyme mix. This enzyme mix can be either mixed directly with a feed, or more conventionally impregnated onto a cereal-based carrier material such as milled wheat, maize or soya flour. Such an impregnated carrier also constitutes an enzyme feed additive in accordance with the present invention.

As an alternative, a cereal-based carrier formed from e.g. milled wheat or maize can

be .impregnated either simultaneously or sequentially with enzymes having the appropriate activities. For example, a milled wheat carrier may be sprayed firstly with a xylanase, secondly with a protease, and optionally finally with a ß-glucanase. The carrier material impregnated with these enzymes also constitutes an enzyme feed additive in accordance with the present invention.

The feed additive of the present invention may be mixed directly with the animal feed, or alternatively mixed with one or more other feed additives such as a vitamin feed additive, a mineral feed additive and an amino acid feed additive. The resulting feed additive including several different types of components can then be mixed in an appropriate amount with the feed.

The feed additive of the invention including the cereal-based carrier is normally mixed in amounts of 0.01-50 g per kilo of feed, more preferably 0.1-10 g/kilo and most preferably about 1 g/kilo.

An alternative way of preparing the enzyme feed additive of the present invention is to construct by recombinant DNA techniques a microorganism which produces the desired enzymes in the desired relative amounts. This can be done for instance by increasing the copy number of the gene encoding protease and/or by using a suitably strong promoter operatively linkded to the polynucleotide encoding the protease. Alternatively or additionally the microorganism strain can be deleted for certain cellulase genes especially endoglucanases.

The enzyme feed additive provided by the present invention may also include other enzymes such as one or more of a xylanase, an alpha.-amylase, a glucoamylase, a pectinase, a mannanase, an .alpha.-galactosidase, a phytase and a lipase. Enzymes having the desired activities may for instance be mixed with the xylanase and protease either before impregnating these on a cereal-based carrier or alternatively such enzymes may be impregnated simultaneously or sequentially on such a cereal-based carrier. The carrier is then in turn mixed with a cereal-based feed to prepare the final feed. It is also possible to formulate the enzyme feed additive as a solution of the individual enzyme activities and then mix this solution with a feed material pre-formed as pellets or as a mash.

It is also possible to include the enzyme feed additive in the animal's diet by incorporating it into a second (and different) feed or drinking water which the animal also has access to. Accordingly, it is not essential that the enzyme mix provided by the present invention is incorporated into the cereal-based feed itself, although such incorporation forms a particularly preferred aspect of the present invention.

The ratio of the units of xylanase activity per g of the feed additive to the units of protease activity per g of the feed additive is preferably 1:0.001-1,000, more preferably 1:0.01-100 and most preferably 1:0.1-10.

As mentioned above, the enzyme mix provided by the present invention is preferably for use as a feed additive in the preparation of a cereal-based feed.

According to a further aspect of the invention, this cereal-based feed comprises at least 25% by weight, more preferably at least 35% by weight, of wheat or maize or a combination of both of these cereals. The feed further comprises a protease in such an amount that the feed includes a protease in such an amount that the feed includes 100-100,000 units of protease activity per kg.

Cereal-based feeds according to the present invention are suitable for animals such as turkeys, geese, ducks, pigs, sheep and cows. The feeds though are particularly suitable for poultry and pigs, and in particular broiler chickens.

One aspect of the invention is a composition for the treatment of a textile that includes proteases of the present invention. The composition can be used to treat for example silk or wool as described in publications such as RE 216,034; EP 134,267; US 4,533,359; and EP 344,259.The serine protease and derivatives thereof as taught herein can be used in a variety of applications where it is desirable to separate phosphorous from phytate. Accordingly, the present invention relates to a method for producing wool or animal hair material with improved properties comprising the steps of

- a. pretreating wool, wool fibres or animal hair material in a process selected from the group consisting of plasma treatment processes and the Delhey process, and
- b. subjecting the pretreated wool or animal hair material to a treatment with a proteolytic enzyme (a protease) in an amount effective for improving the properties.

It is contemplated that the treatment with a proteolytic enzyme can take place prior to the plasma treatment or after the plasma treatment, either in a separate step or e.g. in combination with the scouring or the dyeing of the wool or animal hair material. Further, a surfactant or a softener can be present in the enzyme treatment step, or a separate step wherein the wool or animal hair material is subjected to a softening treatment can be applied.

Published Japanese Patent Application Tokkai Hei 4-327274 discloses a method for a shrink-proofing treatment of e.g. wool fibers by subjecting the fibers to a low-temperature plasma treatment followed by treatment with a shrink-proofing resin, e.g. block-urethane

resin, polyamide epochlorohydrin resin, glyoxalic resin, ethylene-urea resin or acrylate resin, and then a weight reducing treatment with a proteolytic enzyme for obtaining a softening effect. The plasma treatment step is a low-temperature treatment, preferably a corona discharge treatment or a glow discharge treatment.

The Delhey process is described in DE-A-43 32 692 and in J. Delhey: PhD Thesis, RWTH Aachen (1994). In this process the wool is treated in an aqueous solution of hydrogen peroxide in the presence of soluble wolframate, optionally followed by treatment in a solution or dispersion of synthetic polymers, for improving the anti-felting properties of the wool, However, neither does this treatment meet the demands of the end-users.

This low-temperature plasma treatment is carried out by using a gas, preferably a gas selected from the group consisting of air, oxygen, nitrogen, ammonia, helium, or argon. Conventionally, air is used but it may be advantageous to use any of the other indicated gasses.

Preferably, the low-temperature plasma treatment is carried out at a pressure between about 0.1 torr and 5 torr for from about 2 seconds to about 300 seconds, preferably for about 5 seconds to about 100 seconds, more preferably from about 5 seconds to about 30 seconds. The Delhey process is described in J. Delhey: PhD Thesis RWTH Aachen 1994; and in DE-A-43 32 692 and is carried out as follows:

The wool is treated in an aqueous solution of hydrogen peroxide (0.1-35% (w/w), preferably 2-10% (w/w)), in the presence of a 2-60% (w/w), preferably 8-20% (w/w) of a catalyst (preferably $Na_2 WO_4$), and in the presence of a nonionic wetting agent. Preferably, the treatment is carried out at pH 8-11, and room temperature. The treatment time depends on the concentrations of hydrogen peroxide and catalyst, but is preferably 2 minutes or less.

After the oxidative treatment, the wool is rinsed with water.

For removal of residual hydrogen peroxide, and optionally for additional bleaching, the wool may be treated further in acidic solutions of reducing agents (sulphites, phosphites etc.).

The enzyme treatment step is preferably carried out for between about 1 minute and about 120 minutes; preferably at a temperature of between about 20° C. and about 60° C., more preferably between about 30° C. and about 50° C. Alternatively, the wool can be soaked in or padded with an aqueous enzyme solution and then subjected to steaming at a conventional temperature and pressure, typically for about 30 seconds to about 3 minutes.

The proteolytic enzyme treatment is carried out in an acidic or neutral or alkaline medium which may include a buffer.

It may be advantageous to carry out the enzyme treatment step in the presence of one or more conventional anionic, non-ionic or cationic surfactants. An example of a useful nonionic surfactant is Dobanol (from Henkel AG).

Further, the wool or animal hair material may be subjected to an ultrasound treatment, either prior to or simultaneous with the treatment with a proteolytic enzyme. The ultrasound treatment may advantageously be carried out at a temperature of about 50° C. for about 5 minutes.

The amount of proteolytic enzyme used in the enzyme treatment step is preferably between about 0.2 w/w % and about 10 w/w %, based on the weight of the wool or animal hair material.

It is to be understood that, to reduce the number of treatment steps, the enzyme treatment can be carried out during dyeing or scouring of the wool or animal hair material, simply by adding the protease to the dyeing, rinsing or scouring bath.

Preferably, the enzyme treatment is carried out after the plasma treatment but the two treatment steps may also be carried out vice versa.

The softeners conventionally used on wool are usually cationic softeners, either organic cationic softeners or silicone based products, but anionic or non-inoc softeners are also useful. Examples of useful softeners are polyethylene softeners and silicone softeners, i.e. dimethyl polysiloxanes (silicone oils), H-polysiloxanes, silicone elastomers, aminofunctional dimethyl polysiloxanes, aminofunctional silicone elastomers, and epoxyfunctional dimethyl polysiloxanes, and organic cationic softeners, e.g. alkyl quarternary ammonium derivatives.

One aspect of the invention is a composition for the treatment of an animal hide that includes proteases of the present invention. The composition can be used to treat for example animal hide as described in publications such as International Publication, WO 03/00865 (Insect Biotech Co., Taejeon-Si, Korea). In another aspect of the present invention, a method for the processing of hides or skins into leather is described, comprising enzymatice treatment of the hide or skin with the protease of the present invention (International Publication WO 96/11285)

Another aspect of the present invention provides a method for the processing of hides or skins into leather, comprising enzymatic treatment of the hide or skin with a protease described supra herein.

The hides and skins are usually received in the tanneries in the form of salted or dried raw hides or skins. The processing of hides or skins into leather comprises several

different process steps including the steps of soaking, unhairing and bating. These steps constitute the wet processing and are performed in the beamhouse. Enzymatic treatment according to the present invention may take place any time during the manufacture of leather. However, proteases are usually employed during the wet processing, e.g., during soaking, unhairing and/or bating.

In one aspect of the invention, the enzymatic treatment with a protease described supra takes place during the wet processing.

A soaking process of the present invention may be performed at conventional soaking conditions, e.g., at a pH in the range pH 6.0 - 11, in one aspect in the range pH 7.0 –10.0, a temperature: in the range of 20-30 ° C, preferably the range 24-28 ° C, and a reaction time in the range 2-24 hours, preferably the range 4-16 hours, and together with known tensides and preservatives, if needed.

The second phase of the bating step usually commences with the addition of the bate itself. In another specific aspect, the enzymatic treatment takes place during bating. In another aspect, the enzymatic treatment takes place during bating, after the deliming phase. A bating process of the presents invention may be performed at conventional conditions, e.g., at a pH in the range pH 6.0 –9.0, preferably the range pH 6.0 to 8.5, a temperature in the range 20-30° C, preferably the range 25-28°C, and a reaction time in the range 20-90 minutes, preferably the range 40-80 minutes. Processes for the manufacture of leather are well known to the person skilled in the art and have been described in e.g. WO 94/069429 WO901121189 US 3840433, EP-Al -505920, GB-A 2233665 and US 3986926.

Another aspect of the present invention encompasses a bate comprising a protease. A bate is an agent or an enzyme containing preparation comprising the chemically active ingredients for use in beamhouse processes, in particular in the bating step of a process for the manufacture of leather.

In another aspect, the invention provides a bate comprising protease and suitable excipients. In the context of this invention, suitable agents include the auxiliary chemicals known and used in the art, e.g. diluents, emulgators, delimers and carriers. The bate may be formulated as known and described in the art, e.g. as described in GB-A2250289.

The bate of the invention may contain of from 0.00005 to 0.01 g of active protease per g of bate, in one aspect from 0.0002 to 0.004 g of active protease per g of bate. Examples of specific bates recipes are described in WO 96/11285 [EP 0784703B1], Examples 1 and 2.

All patent and literature references cited in the present specification are hereby

incorporated by reference in their entirety.

The invention will be explained further below in the accompanying examples which are provided for illustrative purposes and should not be considered as limitative of the invention.

EXAMPLES:

Unless otherwise indicated, compounds are used as purchased or otherwise obtained without additional purification.

Example 1
Production of 69B4 protease using Gram-positive alkaliphilic bacterium 69B4 protease

1A Methods and materials

Strains

The alkaliphilic micro-organism *Cellulomonas* strain 69B.4, (DSM 16035) used in this study was obtained from Genencor International (Palo Alto, CA). S train 69B4 was stored as frozen glycerol (20% v/v) stocks (at –80 °C).

Origins

A sample of sediment and water from the littoral zone of Lake Bogoria, Kenya at Acacia Camp (Lat. 0° 12'N, Long. 36° 07'E) was collected on 10 October 1988[1]. The water temperature was 33°C, pH 10.5 with a conductivity of 44 mS/cm.

Bacterial strain and culture conditions

Strain 69B4 was isolated at 37°C on an alkaline casein medium containing (g l^{-1})

Glucose (Merck 1.08342)	10
Peptone (Difco 0118)	5
Yeast extract (Difco 0127)	5
K_2HPO_4	1
$MgSO_4.7H_2O$	0.2
NaCl	40

[1] No legal restrictions. This date precedes ratification of the CBD treaty

Na_2CO_3		10
Casein		20
Agar	20	

As described in US Patent 5,401,657 and Duckworth et al. 1996 (Duckworth, A.W., et al, FEMS Microbiology Ecology, 19: 181-191 (1996)

Alkaline cultivation medium

Grant Alkaliphile Medium ("GAM") solution A (g l^{-1})

Glucose (Merck 1.08342)	10
Peptone (Difco 0118)	5
Yeast extract (Difco 0127)	5
K_2HPO_4	1
$MgSO_4.7H_2O$	0.2

Dissolved in 800 ml distilled water and sterilized by autoclaving

GAM solution B (g l^{-1})

NaCl	40
Na_2CO_3	10

Dissolved in 200 ml distilled water and sterilized by autoclaving.

Complete GAM medium is prepared by mixing Solution A (800 ml) with Solution B (200 ml). Solid medium is prepared by the addition of agar (2% w/v).

1B. Growth conditions

From a freshly thawed glycerol vial, the micro-organisms were inoculated using an inoculation needle on Grant Alkaliphile Medium (GAM) described above in agar plates and grown for at least 2 days at 37 °C. One colony was then used to inoculate a 500 ml shake flask containing 100 ml of GAM at pH 10. This flask was then incubated at 37 °C in a rotary shaker at 280 rpm for 1-2 days until good growth (according to visual observation) was obtained. 100 ml of broth culture was subsequently used to inoculate a 7 L fermentor containing 5 liters of GAM. The fermentations were run at 37 °C for 2-3 days in order to obtain maximal production of protease. Fully aerobic conditions were maintained throughout by injecting air, at a rate of 5 L/min, into the region of the impeller, which was rotating at about 500 rpm. The pH was set at pH 10 at the start but was not controlled during the

fermentation.

1.C. Preparation of 69B4 crude enzyme samples

Culture broth was collected from the fermentor, and cells were removed by centrifugation for 30 min at 5000 x *g* at 10 ° C. The resulting supernatant was clarified by depth filtration over Seitz EKS (SeitzSchenk Filtersystems GmbH, Bad Kreuznach, Germany). The resulting sterile culture supernatant was further concentrated approximately 10 times by ultra filtration using an ultrafiltration cassette with a 10kDa cut-off (Pall Omega 10kDa Minisette, Pall Corporation, East Hills, New York, USA). The resulting concentrated crude 69B4 samples were frozen and stored at –20 °C until further use.

1D. Purification

The cell separated culture broth was dialyzed against 20mM (2-(4-morpholino)-ethane sulfonic acid ("MES") ,pH 5.4,1mM $CaCl_2$ using 8K Molecular Weight Cut Off (MWCO) Spectra-Por7 (Spectrum Laboratories, Inc., DominguezRancho , CA, USA) dialysis tubing. The dialysis was performed overnight or until the conductivity of the sample was less than or equal to the conductivity of the MES buffer. The dialyzed enzyme sample was purified using a BioCad VISION(Applied Biosystems, Foster City, CA, USA) with a 10x100mm(7.845mL) POROS High Density Sulfo-propyl (HS) 20 (20micron) cation-exchange column (PerSeptive Biosystems, Framingham, Mass, USA). After loading the enzyme on the previously equilibrated column at 5mL/min, the column was washed at 40mL/min with a pH gradient from 25mM MES, pH 6.2, 1mM $CaCl_2$ to 25mM (N-[2-hydroxyethyl] piperazine-N'-[2-ethane] sulfonic acid [$C_8H_{18}N_2O_4S$, CAS # 7365-45-9]) ("HEPES") pH 8.0,1mM $CaCl_2$ in 25 column volumes. Fractions (8mL) were collected across the run. The pH 8.0 wash step was held for 5 column volumes and then the enzyme was eluted using a gradient (0-100 mM NaCl in the same buffer in 35 column volumes). Protease activity in the fractions were monitored using the pNA assay (sAAPF-pNA assay) (DelMar, E.G., et al 1979). Protease activity which eluted at 40mM NaCl was concentrated and buffer exchanged(using a 5K MWCO VIVA Science 20mL concentrator) into 20mM MES,pH 5.8, 1mMCaCl2. This material was used for further characterization of the enzyme.

Example 2

PCR amplification of a serine protease gene fragment

2A. Degenerate primer design

Based on alignments of published serine protease amino acid sequences (Figs. 11A-B), a range of degenerate primers were designed against conserved structural and catalytic regions. Such regions included those that were highly conserved among the serine proteases, as well as those known to be important for enzyme structure and function.

In one study, protein sequences of nine published serine proteases (Streptogrisin C homologues) were aligned (FIG. 14A-B). The sequences were *Streptomyces griseus* Streptogrisin C (accession no. P52320); alkaline serine protease precursor from *Thermobifida fusca* (accession no. AC23545); alkaline proteinase (EC 3.4.21.-) from *Streptomyces* sp. (accession no. PC2053); alkaline serine proteinase I from *Streptomyces* sp. (accession no. S34672); serine protease from *Streptomyces lividans* (accession no. CAD4208); putative serine protease from *Streptomyces coelicolor* A3(2) (accession no. NP_625129); putative serine protease from *Streptomyces avermitilis* MA-4680 (accession no. NP_822175); serine protease from *Streptomyces lividans* (accession no. CAD42809); putative serine protease precursor from *Streptomyces coelicolor* A3(2) (accession no. NP_628830). It should be noted that all of these sequences are publicly available from GENbank.

Two particular regions were chosen to meet the criteria above, and a forward and a reverse primer designed from the amino acid sequences. The specific amino acid regions used to design the primers are highlighted in black in the protein sequence alignment of Figure 14A-B. Using the genetic code for codon usage, degenerate nucleotide PCR primers were synthesized by MWG-Biotech Ltd. (Ebersberg, Germany)
The degenerated primer sequences are as follows:

forward primer TTGWXCGT_FW: 5' ACNACSGGSTGGCRGTGCGGCAC 3'
reverse primer GDSGGX_RV: 5'-ANGNGCCGCCGGAGTCNCC-3'

As all primers were synthesized in the 5'-3' direction and standard IUB code for mixed base sites was used (e.g. to designate I for A/C/T/G). Degenerate primers TTGWXCGT_FW and GDSGGX_RV successfully amplified a 177 bp region from *Cellulomonas* sp. isolate 69B4 by PCR, as described next.

2B. PCR amplification of a serine protease gene fragment

Cellulomonas sp. isolate 69B4 genomic DNA was used as a template for PCR amplification of putative serine protease gene fragments using the above-described primers. PCR was carried out using High Fidelity Platinum Taq polymerase (Catalog number 11304-102) from Invitrogen Life Technologies (Carlsbad, CA, USA). Conditions were determined by individual experiments, but typically thirty cycles were run in a thermal cycler (MJ Research Inc., Reno, NV, USA). Successful amplification was verified by electrophoresis of the PCR reaction on a 1% agarose gel. A PCR product that was amplified from *Cellulomonas* sp. 69B4 with the primers TTGWXCGT_FW and GDSGGX_RV was purified by gel extraction using the Qiaquick Spin Gel Extraction kit (Catalogue 28704) from Qiagen (Valencia, CA, USA) according to the instructions of the supplier. The purified PCR product was cloned into the commercial pCR2.1TOPO vector System (Invitrogen Life Technologies [Carlsbad, CA, USA]) according to the instructions of the supplier and transformed into competent *E.coli* TOP10 cells. Colonies containing recombinant plasmids were visualized using blue/white selection. For rapid screening of recombinant transformants, plasmid DNA was prepared from cultures of putative positive (white) colonies. DNA was isolated by the plasmid purification kit of Qiagen and sent to Baseclear (Leiden, The Netherlands) for DNA sequencing. One of the clones contained a DNA insert of 177 bp that showed some homology with several streptogrisin-like protease genes of various *Streptomyces* species and also with serine protease genes from other bacterial species. The sequence of this 177 bp fragment is shown in Fig. 13.

2C. Sequence analysis

The sequences were analyzed by BLAST and protein translation sequence tools. BLAST comparison at the nucleotide level showed various levels of homology to published serine protease sequences. Initially, nucleotide sequences were submitted to BLAST (Basic BLAST version 2.0). The program chosen was "blastx", and the database chosen was "nr". Standard/default parameter values were employed. Sequence data for putative *Cellulomonas* 69B4 protease gene fragment was entered as sequence in FASTA format and the query submitted to BLAST to compare the sequences of the present invention to those already in the database. The results returned for the 177 bp fragment a high number of hits for protease genes from e.g. *S. griseus, S. lividans, S. coelicolor, S. albogriseolus, S. platensus, S. fradiae,* and *Streptomyces* sp. It was concluded further inspection of the 177

bp fragment cloned from *Cellulomonas* sp. isolate 69B4 was desired.

Example3

Isolation of a polynucleotide sequence from the genome of Cellulomonas 69B4 encoding a serine protease by Inverse PCR

3A. Inverse PCR on *Cellulomonas* sp. 69B4 genomic DNA to isolate the gene encoding Cellulomonas strain 69B4 protease

Inverse PCR was used to isolate and clone the full-length serine protease gene from *Cellulomonas* sp. 69B4. Based on the DNA sequence of the 177 bp fragment of the *Cellulomonas* protease gene, novel DNA primers were designed:

69B4int_RV1 5'-CGGGGTAGGTGACCGAGGAGTTGAGCGCAGTG-3'
69B4int_FW2 5'-GCTCGCCGGCAACCAGGCCCAGGGCGTCACGTC-3'

Chromosomal DNA of *Cellulomonas* sp. 69B4 was digested with the restriction enzymes *Apa*I, *Bam*HI, *Bss*HII, *Kpn*I, *Nar*I, *Nco*I, *Nhe*I, *Pvu*I, *Sal*I or *Sst*II, purified with the PCR purification kit (Qiagen, Catalogue # 28106) and self-ligated with T4 DNA ligase (Invitrogen Life Technologies) according to the instructions accompanying each. Ligation mixtures were purified with the PCR purification kit (Qiagen) and a PCR was performed with primers 69B4int_RV1 and 69B4int_FW2. PCR on DNA fragments that were digested with *Nco*I and thereafter self-ligated, resulted in a PCR product of approximately 1.3 kb. DNA sequencing (BaseClear, The Netherlands) revealed that this DNA fragment covers the main part of a streptogrisin-like protease gene from *Cellulomonas*. The inventors designated this protease 69B4 protease, and the gene encoding *Cellulomonas* 69B4 protease as the ***asp*** gene. The entire sequence of the *asp* gene was derived by some additional inverse PCR reactions with primer 69B40int_FW2 and a novel DNA primer: 69B4-for4 (5' AAC GGC GGG TTC ATC ACC GCC GGC CAC TGC GGC C 3'). Inverse PCR with these primers on *Nco*I, *Bss*HII, *Apa*I and PvuI digested and self-ligated DNA fragments of genomic DNA of *Cellulomonas* sp. 69B4 resulted in the entire sequence of the *asp* gene (see Figs 1A-C).

3B. Analysis of the sequence of *Cellulomonas* sp. 69B4 protease

A saturated sinapinic acid (3,5-dimethoxy-4-hydroxy cinnamic acid)("SA") solution in

a 1:1 v/v acetonitrile ("ACN")/0.1% formic acid solution was prepared. The resulting mixture was vortexed for 60 seconds and then centrifuged for 20 seconds at 14k rpm. 5µl of the matrix supernatant was transferred to a 0.5 ml Eppendorf tube and 1 µl of a 10 pmole/µl protease (69B4) sample was added to the SA matrix supernatant and vortexed for 5 seconds. 1 µl of the analyte/matrix solution was transferred onto a sample plate and, after being completely dry, analyzed by a Voyager DE-STR (from PerSeptive Biosystems, Framingham, Massachusetts), a matrix assisted laser desorption/ionization – time of flight (MALDI-TOF) mass spectrophotometer with the following settings: Mode of operation: Linear; Extraction mode: Delayed; Polarity: Positive; Accelerating voltage: 25000 V; Extraction delay time: 350 nsec; Acquisition mass range: 4000- 20000 Da; Number of laser shots: 100/spectrum; and Laser intensity: 2351. The resulting spectrum is shown as Fig. 17.

A tryptic map was produced by procedure modified from that as described in Anal. Biochem. 223(1):119-29 (1994) (Christianson, T, et al). The protease solution, containing 10 – 50 µg protease was diluted 1:1 with chilled water in a 1.5 ml microtube. 1.0 N HCl was added to a final concentration of 0.1 N HCl, mixed thoroughly and incubated for 10 minutes on ice. 50% trichloro-acetic acid ("TCA") was added to a final concentration of 10% TCA and mixed. The sample was incubated for 10 minutes on ice, centrifuged for two minutes and the supernatent discarded. 1 ml of cold 90% acetone was added to resuspend the pellet. The resulting sample was then centrifuged for one minute, the supernatent quickly decanted and remaining liquid was removed by vacuum aspiration. The dry pellet was dissolved in 12 µl of 8.0 M urea solution (480 mg urea [Roche, catalog # 1685899) in 0.65 ml of ammonium bicarbonate solution [final concetration of bicarbonate: 0.5 M]) and incubated for 3-5 minutes at 37 degrees C. The solution was slowing diluted with 48 µl of a n-octyl-beta-D-glucopyranoside solution ("o-water") (200 mg of n-octyl-beta-D-glucopyranoside [$C_{14}H_{28}O_6$, f.w. 292.4] in 200 ml of water). 2.0 µl of trypsin (2.5 mg/ml in 1mM HCl) was added and the mixture was incubated for 15 minutes at 37 degrees C. The proteolytic reaction was quenched with 6 µl of 10% trifluoroacetic acid ("TFA"). Insoluble material and bubbles were removed from the sample by centrifugation for one minute.. The tryptic digest was separate by RP-HPLC on 2.1 X 150 mm C-18 column (5µl particle size, 300 angstroms pore size). The elution gradient was formed from 0.1% (v/v) TFA in water and 0.08% (v/v) TFA in acetonitrile at a flow rate of 0.2 ml-min. The column compartment was heated to 50 degrees C. Peptide elution was monitored at 215 nm and data were collected at 215 nm and 280 nm. The samples were then analyzed on a LCQ Advantage

mass spectrometer with a Surveyor HPLC (both from Thermo Finnigan, San Jose, CA USA). The LCQ mass spectrophotometer was run with the following settings: Spray voltage: 4.5kV; Capillary temperature: 225° C. Data processing was performed using TurboSEQUEST and Xcalibur (ThermoFinnigan, San Jose, CA USA). Sequencing of the typtic digest portions was also performed in part by by Argo BioAnalytica (Morris Plains, NJ, USA).

Analysis on the full sequence of the *asp* gene revealed that it encodes a prosequence protease of 495 amino acids (Fig 6, SEQ ID NO.:6). The first 28 amino acids were predicted to form a signal peptide. The mass of the mature chain of 69B4 protease as produced by *Cellulomonas* strain 69B4 has a molecular weight of 18764 (determined by MALDI-TOF). The sequence of the N-terminus of the mature chain was also determined by MALDI-TOF analysis and starts with the sequence FDVIGGNAYTIGGR. It is believed that the 69B4 protease has a unique precursor structure with NH_2- and COOH terminal Pro-sequences, like for instance Aqualysin I of *Thermus aquaticus*. The predicted molecular weight of mature 69B4 protease as shown in Fig. 8 (SEQ ID NO.:8) was 18776.42, which corresponds well with the m.w. of the purified enzyme with proteolytic activity isolated from *Cellulomonas* sp. 69B4 (18764). The prediction of the COOH terminal pro-sequence in 69B4 protease was also based on an alignment of the 69B4 protease with Aqualysin I of *T. aquaticus* (Fig. 16). The sequence of three internal peptides of the purified enzyme from *Cellulomonas* sp. 69B4 having proteolytic activity was determined by MALDI-TOF analysis. All three peptides were also identified in the translation product of the isolated *asp* gene, confirming the identification of the correct protease gene (see Fig. 1A-C).

3C. Percentage identity comparison between bacterial serine proteases

The deduced polypeptide product of the *asp* gene (mature chain) was used for homology analysis with other serine proteases using the BLAST program and settings as described in Example 2C. The analysis showed identities of from about 44 - 48% (See Table 1, below) and, together with analysis of the translated sequence, provided evidence that the *asp* gene encodes a protease having less than 50% sequence identity with the mature chains of Streptogrisin-like serine proteases. An alignment of Asp with Streptogrisin A, Streptogrisin B, Streptogrisin C, Streptogrisin D of *Streptomyces griseus* is shown in Fig. 15.

Tabl 1. Percentage Identity: Comparison betwe n *Cellulomonas* sp. 69B4 protease encoded by asp and other serine proteases (identity betw en the mature chains)

	Streptogrisin A *S. griseus*	Streptogrisin B S. griseus	Streptogrisin C S. griseus	Streptogrisin D *S. griseus*	Alpha-lytic endopeptidase *Lysobacter enzymogenes*
Asp protease *Cellulomonas* sp. Isolate 69B4	48%	45%	47%	46%	44%

Example 4

Protease Production in *Streptomyces lividans*

The inventors also constructed a plasmid comprising a polypeptide encoding a polypeptide having proteolytic activity and used such vector to transform a *Streptomyces lividans* host cell The transformation is more fully described in US Patent No. 6,287,839 (WO 02/50245), which is herein expressly incorporated by reference.

This example describes the construction of a vector, e.g., a plasmid, comprising a polynucleotide encoding a protease of this invention and used to transform *Streptomyces lividans*. The final plasmid vector is referred to as pSEG69B4T.

The construction of pSEG69B4T made use of one pSEGCT plasmid vector described in International Publication WO 02/50245.

A Glucose Isomerase ("GI") promoter operably linked to the structural gene encoding the 69B4 protease was used to drive the expression of the protease. A fusion between the GI-promoter and the 69B4 signal-sequence, N-terminal prosequence and mature sequence was constructed by fusion-PCR techniques, as a *XbaI-BamHI* fragment. The fragment was ligated into plasmid pSEGCT digested with *XbaI* and *BamHI*, resulting in plasmid pSEG69B4T. (see Fig. 18)

It is contemplated by the inventor that other expression vectors can be prepared using different promoters and / or signal sequences combined with different combinations of 69B4 protease: + / - N terminal and C terminal prosequence in the pSEGCT backbone (vector).

The host *Streptomyces lividans* TK23 was transformed with plasmid vector pSEG69B4T using the protoplast method described in Hopwood, *et al.*,. GENETIC MANIPULATION OF STREPTOMYCES, A LABORATORY MANUAL. The John Innes Foundation, Norwich, United Kingdom (1985).

The transformed culture was expanded to provide two fermentation cultures. At various time points, samples of the fermentation broths were removed for analysis. For the

purposes of this experiment, a skin milk procedure was used to confirm successful cloning. 30 µl of the shake flask supernatent was spotted in punched out holes in skim milk agar plates and incubated at 37 C. The incubated plates were visually reviewed after overnight incubation for the presence of halos. For purposes of this experiment, the samples were also assayed for protease activity and for molecular weight (SDS-PAGE). At the end of the fermentation run full length protease was observed by SDS-PAGE.

A sample of the fermentation broth was assayed as follows: 10µl of the diluted supernatent was taken and added to 190 µl AAPF substrate solution (conc. 1 mg/ml, in 0.1 M Tris/0.005% Tween, pH 8.6.

The rate of increase in absorbance at 410 nm due to release of *p*-nitroaniline was monitored (25°). The assay results of the fermentation broth of 3 clones (X, Y, W) are shown in Table 2, below.

Table 2

CLONE	Total mOD/min	Ppm
V	23296	127
W	19224	105
Y	13171	72

From these results it can be seen that the polynucleotide encoding a polypeptide having proteolytic activity was expressed in *Streptomyces lividans*.

EXAMPLE 5

Determination of Cleaning Activity

An incubator (Innova 4330 Model Incubator, New Brunswick Scientific Co., Edison, NJ, USA) was pre-warmed at 40°C for "European" conditions and for 20° C for Japanese conditions for 60 minutes. Blood-Milk-Ink swatches (EMPA 116) were obtained from the Swiss Federal Laboratories for Material Testing and from Center for Test Materials ("CFT", Vlaardingen, Netherlands) and were modified by exposure to 0.03 % hydrogen peroxide for 30 minutes at 60° C., then dried. Circles of 1/4" diameter were cut from the dried swatches and placed vertically, one per well, in a 96 well microplate. Protease samples as described in Example 1, were diluted into 10 mM NaCl, 0.005% Tween 80 (polyoxyethylene sorbitan

mono-oleate [Sigma P-1754]) to obtain the desired concentration of 10 ppm (protein). 1 gram per liter TIDE laundry detergent (Procter & Gamble, Cinncinati, OH) without bleach and enzyme was prepared in deionized water, and a concentrated stock of calcium and magnesium was added to result in a final water hardness value of 6 grains per gallon [North American wash conditions]. 7.6 gram per liter ARIEL REGULAR laundry detergent (Procter & Gamble, Cinncinati, OH) without bleach and enzyme was prepared in deionized water, and a concentrated stock of calcium and magnesium was added to result in a final water hardness value of 15 grains per gallon [European wash conditions.] In another set of samples, 0.67 gram per liter PURE CLEAN laundry detergent (Procter & Gamble, Cinncinati, OH) without bleach and enzyme was prepared in deionized water, and a concentrated stock of calcium and magnesium was added to result in a final water hardness value of 3 grains per gallon [Japanese wash conditions.] In another detergent composition, 0.66 gram per liter TIDE OPAL laundry detergent without bleach and enzyme was prepared in deionized water, and a concentrated stock of calcium and magnesium was added to result in a final water hardness value of 3 grains per gallon [Japanese wash conditions with North American detergent formulation]. The detergent was allowed to mix for 15 minutes and was then filtered through a 0.2 micron cellulose acetate filter. A 190 ul of the respective detergent solution was then added to the appropriate wells of the microplate. A 10 ul of the enzyme was added to the filtered detergent in order to obtain a final concentration 0.25-3.0 ppm (micrograms per milliliter). Total volume should have been 200 ul. The microplate was then sealed to prevent leakage, placed in a holder on an incubator/shaker set to 20°C and 350/400 RPM and allowed to shake for one hour. The plate was then removed from the incubator/shaker and an aliquot of 100ul of the reaction liquor was removed from each well, and placed on a fresh Costar microtiter plate (Corning Incorporated, Corning, NY, USA). The absorbance at 405 nm wavelength was read for each aliquot on a Microtiter plate reader (SpectraMax 340, Molecular Devices Corp., Sunnyvale, CA, USA) and reported. Detergent composition and incubation conditions in the microswatch assay are set forth in Table 3.

Table 3

Geography	Detergent	Water Hardness	Enzyme dosage	Temperature	Swatch

Powder detergent					
European	Ariel Regular 7.6 g/l	15 gpg Ca/Mg=4/1	0.25 – 3.0 ppm	40°	Superfix
North American	Tide Opal 1.0 g/l	6 gpg Ca/Mg=3/1	0.25 – 3.0 ppm	20°	3K
Japanese	Pure Clean 0.66 g/l	3 gpg Ca/Mg=3/1	0.25 – 3.0 ppm	20°	3K
Japanese	Tide Opal	3 gpg	0.25 – 3.0 ppm	20°	3K
(pseudo)	0.66 g/l	Ca/Mg=3/1			
Liquid detergent (1.5 ml/L)	Liquid-Tide	6 gpg	0.25 – 3.0 ppm	20°	3K

The results of the dose response curves depicting absorbance at 405 nm as a function of concentration [ppm in well], for PURAFECT (Genencor International, Palo Alto, CA, USA), OPTIMASE (Genencor International, Palo Alto, CA, USA), RELASE [GG36-var.1], and the Protease of the present invention are shown in Figures 19 through 23. Under North American conditions, in liquid TIDE (Fig. 19), the 69B4 protease showed enhanced cleaning performance as compared to PURAFECT, RELASE and OPTIMASE proteases under the same conditions. Under Japanese conditions, in TIDE OPAL powder (0.66 g/l) 69B4 showed enhanced or the same cleaning performance as compared to PURAFECT, RELASE and OPTIMASE proteases under the same conditions (Fig. 20). Under European conditions, in ARIEL REGULAR powder detergent (Fig. 21), the 69B4 protease showed enhanced cleaning performance as compared to PURAFECT, RELASE and OPTIMASE proteases under the same conditions. In both 69B4 and OPTIMASE showed 2 to 10 times the absorbance at 405 nm as compared to PURAFECT and RELASE. Under Japanese conditions, in PURE CLEAN powder detergent (Fig. 22), the 69B4 protease showed enhanced and comparative cleaning performance as compared to PURAFECT, RELASE and OPTIMASE proteases under the same conditions. Under North American conditions, in TIDAL OPAL powder detergent (Fig. 23), the 69B4 protease showed enhanced or comparative cleaning performance as compared to PURAFECT, RELASE and OPTIMASE proteases under the same conditions.

Example 6

Determination of Oxidative stability.

The resistance to oxidation of *Cellulomonas* 69B4 protease as compared to a BPN'-variant protease (provided by Genencor International), a GG36 variant protease (provided by Genencor International) and PURAFECT (provided by Genencor International) was assayed by incubating a sample of the protease with 0.1 M H_2O_2. The BPN'variant is described in US RE 34,606. The GG36 variant is described in US Pat. Nos. 5,955,340 and/or 5,700,676. A volume of 2.0 ml 0.1 M Borate buffer (45.4 gm NaB_4O_7 10 H_2O), pH 9.45 containing 0.1 M H_2O_2 and 100 ppm protease was incubated at 25°C for 20 minutes and assayed for enzyme activity as follows: 50 µl of the incubation mixture was combined with 950 µl 0.1 M Tris buffer, pH 8.6 and a sample from 10 µl was taken and added to 990 µl AAPF substrate solution, conc. 1 mg/ml, in 0.1 M Tris / 0.005% Tween, pH 8.6. The rate of increase in absorbance at 410 nm is due to release of *p*-nitroaniline was monitored. The results of 4 proteases are shown in Fig 24. Protease 69B4 showed greatly enhanced stability under oxidative conditions relative to the subtilisin proteases.

Example 7

Determination of Chelate stability.

The resistance to the presence of a chelator of 69B4 protease was assayed by incubating with 10 mM EDTA in 50 mM Tris, pH 8.2 . A volume of 2.0 ml 50 mM Tris buffer, pH 8.2, containing 10 mM EDTA and 100 ppm protease was incubated at 45°C for 100 minutes and assayed for enzyme activity as follows: 50 µl of the incubation mixture was combined with 950 µl 0.1 M Tris buffer, pH 8.6 and a sample from 10 µl was taken and added to 990 µl AAPF substrate solution, conc. 1 mg/ml, in 0.1 M Tris / 0.005% Tween, pH 8.6
The rate of increase in absorbance at 410 nm due to release of p-nitroaniline was monitored. The results of 4 proteases are shown in fig 25. Protease 69B4 showed greatly enhanced stability in the presence of a chelator than a BPN' variant, PURAFECT, or a GG36 variant.

Example 8

Determination of Thermal stability.

Cellulomonas 69B4 protease was tested for resistance to thermal inactivation in solution. The thermal inactivation was performed by incubating a volume of 2.0 ml 50 mM Tris buffer, pH 8.0, containing 100 ppm protease at 45°C for 300 minutes and assayed for enzyme activity as follows: 50 µl of the incubation mixture was combined with 950 µl 0.1 M Tris buffer, pH 8.6 and a sample from 10 µl was taken and added to 990 µl AAPF substrate solution, conc. 1 mg/ml, in 0.1 M Tris / 0.005% Tween, pH 8.6. The rate of increase in absorbance at 410 nm due to release of p-nitroaniline was monitored. The results of 4 proteases are shown in Figure 26. Protease 69B4 showed enhanced or comparative thermal stability at 45 degrees centigrade than a BPN' variant, PURAFECT, or a GG36 variant.

Example 9

Determination of Thermostabilty of 69B4 Protease (2)

Another thermostability study of the 69B4 protease was performed using a temperature gradient between 57°- 62 °C.
The thermal inactivation (using a Thermocycler –MTP plate [DNA Engine Tetad, MJ Research Inc., Waltham, MA, USA) was performed by incubating a volume of 180µl 100 mM Tris buffer, pH 8.6, containing 1 mM $CaCl_2$ and 5 ppm protease for 60 minutes and assayed for enzyme activity as follows: 10 µl was taken and added to 190 µl AAPF substrate solution, conc. 1 mg/ml, in 0.1 M Tris / 0.005% Tween, pH 8.6
The rate of increase in absorbance at 410 nm due to release of *p*-nitroaniline was monitored (25°). The results of 4 proteases are shown in Figure 27.

Example 10

Determination of pH profile of *Cellulomonas* 69B4 protease on DMC substrate.

The *Cellulomonas* 69B4 protease of the present invention, isolated and purifed by methods described herein and three currently used subtilisin proteases (PURAFECT, BPN'-var.1, GG36-var.1)

supplied by Genencor International (Palo Alto, CA, USA), were analyzed for their ability to hydrolyze a commercial synthetic substrate, di-methyl casein ("DMC"/ Sigma C-9801) in the pH range from 4 to 12.

. A 5 mg/ml DMC substrate solution was prepared in the appropriate buffer (5 mg/ml DMC, 0.005% (w/w) Tween 80® (polyoxyethylene sorbitan mono-oleate, Sigma P-1754)). The appropriate DMC buffers were composed as follows: 40 mM MES for pH 4 and 5 ; 40 mM HEPES for pH 6 and 7, 40 mM TRIS for pH 8 and 9; and 40 mM Carbonate for pH 10, 11 and 12.

180 µl of each pH-substrate solution was transferred into 96 well microtiter plate and were pre-incubated at 37 ° C for twenty minutes prior to enzyme addition. The respective enzyme solutions (BPN'-var.1; GG36-var.1; PURAFECT; 69B4 protease) were prepared containing about 25 ppm and 20 µl of these enzyme solutions were pipetted into the substrate containing wells in order to achieve a 2.5 ppm final enzyme concentration in each well. The 96 well plate containing enzyme-substrate mixtures was incubated at 37 ° C and 300 rpm for one hour in an IKS-Multitron incubator/shaker.

Detection method: A 2,4,6-trinitrobenzene sulfonate ("TNBS") color reaction method was used to determine the amount of peptides and amino acids release from DMC substrate. The free amino groups (of the peptides and amino acids) react with 2,4,6-trinitro-benzene sulfonic acid to form a yellow colored complex.
The absorbance is measured at 405 nm in a SpectraMax 250 MTP Reader.

TNBS assay: A 1 mg/ml solution of TNBS (5% 2,4,6 Trinitrobenzene sulfonic acid/Sigma-P2297) was prepared in reagent buffer A (2.4 g NaOH, 45.4 g $Na_2B_4O_7.10H_2O$ dissolved by heating in 1000ml). 60 µl per well was aliquoted into a 96-well plate and 10 µl of the incubation mixture described above was added to each well and mixed for 20 minutes at room temperature. 200 µl of reagens B (70.4 g $NaH_2PO_4 \cdot H_2O$ and 1.2 g Na_2SO_3 in 2000 ml) was added to each well and mixed to stop the reaction. The absorbance at 405 nm was measured in a SpectraMax 250 MTP Reader. The absorbance value was corrected for a blank (without enzyme).
The data in the Table 4 below show the comparative ability of the 69B4 protease to hydrolyze such substrate versus proteases from a known mutant variants (BPN' var. 1 and

GG36 var.1).

As can be seen in the Figure 28, the serine protease of the present invention showed comparative or increased hydrolysis of DMC substrate with an optimal DMC-hydrolysis activity over a broad pH range from 7 to 12.

Results

Table 4	TNBS response (OD405 nm)								
	pH4	pH5	pH6	pH7	pH8	pH9	pH10	pH11	pH12
BPN'var.1	0.095	0.174	0.482	0.749	0.813	0.847	0.730	0.683	0.590
GG36 var.1	0.228	0.172	0.499	0.740	0.958	1.062	1.068	1.175	1.136
Purafect	0.042	0.202	0.545	0.783	0.956	1.130	1.102	1.188	1.174
69B4	0.252	0.218	0.575	0.742	0.803	0.965	0.762	0.741	0.729

Example 11

pH Stability of 69B4 protease.

Determination of the low pH stability of *Cellulomonas* 69B4 protease.

The *Cellulomonas* 69B4 protease of the present invention, isolated and purifed by methods described herein and two currently used subtilisin proteases (PURAFECT, BPN'-var.1 [US RE 34,606]) supplied by Genencor International (Palo Alto, CA, USA), were analyzed for their stability in the pH range from 3 to 6.

. The respective enzyme solutions (BPN'-var.1, PURAFECT, 69B4 protease) were prepared containing 90 ppm protease in 0.1 M Citrate buffer, pH 3, 4, 5 and 6

10 ml tubes containing 1 ml of buffered enzyme solutions were placed for 60 minutes in a GFL 1083 water bath set at 25°C, 35°C and 45°C respectively.

AAPF activity was determined for each enzyme sample at time 0 and 60 minutes as described in Example 10. The remaining enzyme activity was calculated and the results are provided in Table 5 below (Figures 29-32).

The data in the Table show the exceptional stability of the 69B4 protease at pH 3, 4, 5, and 6 between 25°C and 45°C compared to BPN' var. 1 and PURAFECT.

Results

Table 5	BPN' Var. 1			PURAFECT			69B4		
	25°	35°	45°	25°	35°	45°	25°	35°	45°
pH 3	39	1	0	42	2	0	97	109	95
pH 4	92	35	1	55	7	0	106	105	102
pH 5	112	82	12	95	68	8	114	115	106
pH 6	113	99	59	104	96	63	95	104	104

EXAMPLE 12

Liquid dishwashing Composition

Liquid fabric cleaning composition of particular utility under Japanese machine wash conditions could be prepared in accordance with the invention.

AE2.5S	15.00
AS	5.50
N-Cocoyl N-methyl glucamine	5.50
Nonionic surfactant	4.50
Citric acid	3.00
Fatty acid	5.00
Base	0.97
Monoethanolamine	5.10
1,2-Propanediol	7.44
EtOH	5.50
HXS	1.90
Boric Acid	3.50
Ethoxylated tetraethylenepentaimine	3.00
SRP	0.30
Protease	0.069
Amylase	0.06
Cellulase	0.08
Lipase	0.18
Brightener	0.10
Minors/inerts to 100%	

EXAMPLE 13

Liquid Fabric Cleaning Composition of particular utility under Japanese machine wash conditions and for fine fabrics could be prepared in accordance with the invention.

Component	Amount (%)
AE2.5S	2.16
AS	3.30
N-Cocoyl N-methyl glucamine	1.10
Nonionic surfactant	10.00
Citric acid	0.40
Fatty acid	0.70
Base	0.85
Monoethanolamine	1.01
1,2-Propanediol	1.92
EtOH	0.24
HXS	2.09
Protease.sup.1	0.01
Amylase	0.06
Minors/inerts to 100%	

EXAMPLE 14

Liquid Dishwashing Compositions (especially suitable under Japanese conditions) can be prepared in accordance with the present invention.

Component	A	B
AE1.4S	24.69	24.69
N-cocoyl N-methyl glucamine	3.09	3.09
Amine oxide	2.06	2.06
Betaine	2.06	2.06
Nonionic surfactant	4.11	4.11
Hydrotrope	4.47	4.47

Magnesium	0.49	0.49
Ethanol	7.2	7.2
LemonEase	0.45	0.45
Geraniol/BHT	—	0.60/0.02
Amylase	0.03	0.005
Protease	0.01	0.43
Balance to 100%		

EXAMPLE 15

Granular Fabric Cleaning compositions can be prepared in accordance with the invention.

Component	A	B	C	D
Protease1	0.10	0.20	0.03	0.05
Protease2			0.2	0.15
C13 linear alkyl benzene sulfonate	22.00	22.00	22.00	22.00
Phospate (as sodium tripolyphosphate)	23.00	23.00	23.00	23.00
Sodium carbonate	23.00	23.00	23.00	23.00
Sodium silicate	14.00	14.00	14.00	14.00
Zeolite	8.20	8.20	8.20	8.20
Chelant (diethylaenetriamine-petaacetic acid)	0.40	0.40	0.40	0.40
Sodium sulfate	5.50	5.50	5.50	5.50
Water	Balance to 100%			

EXAMPLE !6

A Granular Fabric Cleaning composition can be prepared in accordance with the invention.

	A	B	C	D
Protease1	0.10	0.20	0.30	0.05
Protease2			0.2	0.1
C12 alkyl benzene sulfonate	12.00	12.00	12.00	12.00
Zeolite A (1-10 micrometer)	26.00	26.00	26.00	26.00
C12-C14 secondary (2,3) alkyl sulfate, Na salt	5.00	5.00	5.00	5.00
Sodium citrate	5.00	5.00	5.00	5.00
Optical brightenere	0.10	0.10	0.10	0.10
Sodium sulfate	17.00	17.00	17.00	17.00
Fillers, water, minors	Balance to 100%			

EXAMPLES 17

The following granular laundry detergent compositions 37 A-C are of particular utility under European machine wash conditions can be prepared in accordance with the invention:

component	A	B	C
LAS	7.0	5.61	4.76
TAS			1.57
C45AS	6.0	2.24	3.89
C25E25	1.0	0.76	1.18
C45E7			2.0
C25E3	4.0	5.5	
QAS	0.8	2.0	2.0
STPP			
Zeolite	25.0	19.5	19.5
Citric acid	2.0	2.0	2.0
NaSKS-6	8.0	10.6	10.6
Carbonate I	8.0	10.0	8.6
MA/AA	1.0	2.6	1.6
CMC	0.5	0.4	0.4
PB4		12.7	
Percarbonate			19.7
TAED		3.1	5.0
Citrate	7.0		
DTPMP	0.25	0.2	0.3
HEDP	0.3	0.3	0.3
QEA 1	0.9	1.2	1.0
Protease1	0.02	0.05	0.035
Lipase	0.15	0.25	0.15
Cellulase	0.28	0.28	0.28
Amylase	0.4	0.7	0.3
PVPI/PVNO	0.4		0.1
Photoactivated bleach (ppm)	15 ppm	27 ppm	27 ppm
Brightener 1	0.08	0.19	0.19
Brightener 2		0.04	0.04
Perfume	0.3	0.3	0.3
Efferevescent granules (malic acid 40%, sodium bicarbonate 40%, sodium carbonate 20%)	15	15	5

Silicon antifoam	0.5	2.4	2.4
Minors/inerts to 100%	Balance to 100%		

EXAMPLE 18

Four groups of Nicholas male turkey poults could be fed the maize-based starter feed described in Table 8 in the form of a mash up to 21 days of age.

TABLE 8

Ingredient Amount	(wt. %)
Maize	36.65
Soybean meal (45.6% CP)	55.4
Animal-vegetable fat	3.2
Dicalcium phosphate	2.3
Limestone	1.5
Mineral premix	0.3
Vitamin premix	0.3
Sodium chloride	0.15
DL methionine	0.2

The first control group could be fed the feed of Table 8 unsupplemented. The feeds of three other test groups could be supplemented respectively with 2,000 units/kg, 4,000 units/kg and 6,000 units/kg of a *Cellulomonas* protease.

Of course, it should be understood that a wide range of changes and modifications can be made to the preferred embodiment described above. It is therefore intended that the foregoing detailed description be understood in the context of any claims, including all equivalents, which are intended to define the scope of this invention.

ABSTRACT

A novel protease comprising at least 70% amino acid sequence identity to SEQ ID NO.:8 is provided. An isolated polynucleotide encoding a protease comprising an amino acid sequence with at least 70% sequence identity to SEQ ID NO.:8 is provided. The protease and/or polynucleotide may be derived from a cellulomonas spp, for example Cellulomonas strain 69B4 (DSM 16035). Also provided for is a method of producing an enzyme having proteolytic activity comprising the steps of transforming a host cell with an expression vector comprising a polynucleotide having at least 70% sequence identity to SEQ ID NO. 4; cultivating the transformed host cell under conditions suitable for the host cell to produce the protease and recovering the protease. The use of the expressed serine protease protein in cleaning compositions and in feed as a supplement are also described.

FIG 1A

```
1     GCGCGCTGCG CCCACGACGA CGCCGTCCGC CGTTCGCCGG CGTACCTGCG TTGGCTCACC
      CGCGCGACGC GGGTGCTGCT GCGGCAGGCG GCAAGCGGCC GCATGGACGC AACCGAGTGG

61    ACCCACCAGA TCGACCTCCA TAACGAGGCC GTATGACCAG AAAGGGATCT GCCACCGCCC
      TGGGTGGTCT AGCTGGAGGT ATTGCTCCGG CATACTGGTC TTTCCCTAGA CGGTGGCGGG

121   ACCAGCACGC TCCTAACCTC CGAGCACCGG CGACCGCCGG GTGCGATGAA AGGGACGAAC
      TGGTCGTGCG AGGATTGGAG GCTCGTGGCC GCTGGCGGCC CACGCTACTT TCCCTGCTTG

181   CGAGATGACA CCACGCACAG TCACGCGGGC CCTGGCCGTG GCCACCGCAG CCGCCACACT
      GCTCTACTGT GGTGCGTGTC AGTGCGCCCG GGACCGGCAC CGGTGGCGTC GGCGGTGTGA

241   CCTGGCAGGC GGCATGGCCG CCCAGGCCAA CGAGCCCGCA CCACCCGGGA GCGCGAGCGC
      GGACCGTCCG CCGTACCGGC GGGTCCGGTT GCTCGGGCGT GGTGGGCCCT CGCGCTCGCG

301   ACCGCCACGC CTGGCCGAGA AGCTCGACCC CGACCTCCTC GAGGCCATGG AGCGCGACCT
      TGGCGGTGCG GACCGGCTCT TCGAGCTGGG GCTGGAGGAG CTCCGGTACC TCGCGCTGGA

361   GGGCCTCGAC GCGGAGGAAG CCGCCGCCAC CCTGGCGTTC CAGCACGACG CAGCCGAGAC
      CCCGGAGCTG CGCCTCCTTC GGCGGCGGTG GGACCGCAAG GTCGTGCTGC GTCGGCTCTG

421   CGGCGAGGCC CTCGCCGAAG AGCTCGACGA GGACTTCGCC GGCACCTGGG TCGAGGACGA
      GCCGCTCCGG GAGCGGCTTC TCGAGCTGCT CCTGAAGCGG CCGTGGACCC AGCTCCTGCT

481   CGTCCTGTAC GTCGCCACCA CCGACGAGGA CGCCGTCGAG GAGGTCGAGG GCGAAGGCGC
      GCAGGACATG CAGCGGTGGT GGCTGCTCCT GCGGCAGCTC CTCCAGCTCC CGCTTCCGCG

541   CACGGCCGTC ACCGTCGAGC ACTCCCTGGC CGACCTCGAG GCCTGGAAGA CCGTCCTCGA
      GTGCCGGCAG TGGCAGCTCG TGAGGGACCG GCTGGAGCTC CGGACCTTCT GGCAGGAGCT

601   CGCCGCCCTC GAGGGCCACG ACGACGTGCC CACCTGGTAC GTCGACGTCC CGACCAACAG
      GCGGCGGGAG CTCCCGGTGC TGCTGCACGG GTGGACCATG CAGCTGCAGG GCTGGTTGTC

661   CGTCGTCGTC GCCGTCAAGG CCGGAGCCCA GGACGTCGCC GCCGGCCTCG TCGAAGGTGC
      GCAGCAGCAG CGGCAGTTCC GGCCTCGGGT CCTGCAGCGG CGGCCGGAGC AGCTTCCACG

721   CGACGTCCCG TCCGACGCCG TGACCTTCGT CGAGACCGAC GAGACCCCGC GGACCATGTT
      GCTGCAGGGC AGGCTGCGGC ACTGGAAGCA GCTCTGGCTG CTCTGGGGCG CCTGGTACAA

781   CGACGTGATC GGCGGCAACG CCTACACCAT CGGGGGGCGC AGCCGCTGCT CGATCGGGTT
      GCTGCACTAG CCGCCGTTGC GGATGTGGTA GCCCCCCGCG TCGGCGACGA GCTAGCCCAA

841   CGCGGTCAAC GGCGGGTTCA TCACCGCCGG CCACTGCGGC CGCACCGGCG CCACCACCGC
      GCGCCAGTTG CCGCCCAAGT AGTGGCGGCC GGTGACGCCG GCGTGGCCGC GGTGGTGGCG
```

FIG 1B

```
 901    CAACCCCACC GGGACCTTCG CCGGGTCCAG CTTCCCGGGC AACGACTACG CGTTCGTCCG
        GTTGGGGTGG CCCTGGAAGC GGCCCAGGTC GAAGGGCCCG TTGCTGATGC GCAAGCAGGC

 961    TACCGGGGCC GGCGTGAACC TGCTGGCCCA GGTCAACAAC TACTCCGGTG GCCGCGTCCA
        ATGGCCCCGG CCGCACTTGG ACGACCGGGT CCAGTTGTTG ATGAGGCCAC CGGCGCAGGT

1021    GGTCGCCGGG CACACCGCGG CCCCCGTCGG CTCGGCCGTG TGCCGGTCCG GGTCGACCAC
        CCAGCGGCCC GTGTGGCGCC GGGGGCAGCC GAGCCGGCAC ACGGCCAGGC CCAGCTGGTG

1081    CGGGTGGCAC TGCGGCACCA TCACTGCGCT CAACTCCTCG GTCACCTACC CCGAGGGCAC
        GCCCACCGTG ACGCCGTGGT AGTGACGCGA GTTGAGGAGC CAGTGGATGG GGCTCCCGTG

1141    CGTCCGCGGC CTGATCCGCA CCACCGTCTG CGCCGAGCCC GGCGACTCCG GTGGCTCGCT
        GCAGGCGCCG GACTAGGCGT GGTGGCAGAC GCGGCTCGGG CCGCTGAGGC CACCGAGCGA

1201    GCTCGCCGGC AACCAGGCCC AGGGCGTCAC GTCCGGCGGC TCCGGCAACT GCCGCACCGG
        CGAGCGGCCG TTGGTCCGGG TCCCGCAGTG CAGGCCGCCG AGGCCGTTGA CGGCGTGGCC

1261    TGGCACCACG TTCTTCCAGC CGGTCAACCC CATCCTCCAG GCGTACGGCC TGAGGATGAT
        ACCGTGGTGC AAGAAGGTCG GCCAGTTGGG GTAGGAGGTC CGCATGCCGG ACTCCTACTA

1321    CACCACGGAC TCGGGCAGCA GCCCGGCCCC TGCACCGACC TCCTGCACCG GCTACGCCCG
        GTGGTGCCTG AGCCCGTCGT CGGGCCGGGG ACGTGGCTGG AGGACGTGGC CGATGCGGGC

1381    CACCTTCACC GGGACCCTCG CGGCCGGCCG GGCCGCCGCC CAGCCCAACG GGTCCTACGT
        GTGGAAGTGG CCCTGGGAGC GCCGGCCGGC CCGGCGGCGG GTCGGGTTGC CCAGGATGCA

1441    GCAGGTCAAC CGGTCCGGGA CCCACAGCGT GTGCCTCAAC GGGCCCTCCG GTGCGGACTT
        CGTCCAGTTG GCCAGGCCCT GGGTGTCGCA CACGGAGTTG CCCGGGAGGC CACGCCTGAA

1501    CGACCTCTAC GTGCAGCGCT GGAACGGCAG CTCCTGGGTG ACCGTCGCCC AGAGCACCTC
        GCTGGAGATG CACGTCGCGA CCTTGCCGTC GAGGACCCAC TGGCAGCGGG TCTCGTGGAG

1561    CCCCGGCTCC AACGAGACCA TCACCTACCG CGGCAACGCC GGCTACTACC GCTACGTGGT
        GGGGCCGAGG TTGCTCTGGT AGTGGATGGC GCCGTTGCGG CCGATGATGG CGATGCACCA

1621    CAACGCCGCG TCCGGCTCCG GTGCCTACAC CATGGGGCTC ACCCTCCCCT GACGTAGCGC
        GTTGCGGCGC AGGCCGAGGC CACGGATGTG GTACCCCGAG TGGGAGGGGA CTGCATCGCG
```

FIG 2A

```
   1    ATGACACCAC GCACAGTCAC GCGGGCCCTG GCCGTGGCCA CCGCAGCCGC CACACTCCTG
        TACTGTGGTG CGTGTCAGTG CGCCCGGGAC CGGCACCGGT GGCGTCGGCG GTGTGAGGAC
                                  .85
  61    GCAGGCGGCA TGGCCGCCCA GGCCAACGAG CCCGCACCAC CCGGGAGCGC GAGCGCACCG
        CGTCCGCCGT ACCGGCGGGT CCGGTTGCTC GGGCGTGGTG GGCCCTCGCG CTCGCGTGGC

 121    CCACGCCTGG CCGAGAAGCT CGACCCCGAC CTCCTCGAGG CCATGGAGCG CGACCTGGGC
        GGTGCGGACC GGCTCTTCGA GCTGGGGCTG GAGGAGCTCC GGTACCTCGC GCTGGACCCG

 181    CTCGACGCGG AGGAAGCCGC CGCCACCCTG GCGTTCCAGC ACGACGCAGC CGAGACCGGC
        GAGCTGCGCC TCCTTCGGCG GCGGTGGGAC CGCAAGGTCG TGCTGCGTCG GCTCTGGCCG

 241    GAGGCCCTCG CCGAAGAGCT CGACGAGGAC TTCGCCGGCA CCTGGGTCGA GGACGACGTC
        CTCCGGGAGC GGCTTCTCGA GCTGCTCCTG AAGCGGCCGT GGACCCAGCT CCTGCTGCAG

 301    CTGTACGTCG CCACCACCGA CGAGGACGCC GTCGAGGAGG TCGAGGGCGA AGGCGCCACG
        GACATGCAGC GGTGGTGGCT GCTCCTGCGG CAGCTCCTCC AGCTCCCGCT TCCGCGGTGC

 361    GCCGTCACCG TCGAGCACTC CCTGGCCGAC CTCGAGGCCT GGAAGACCGT CCTCGACGCC
        CGGCAGTGGC AGCTCGTGAG GGACCGGCTG GAGCTCCGGA CCTTCTGGCA GGAGCTGCGG

 421    GCCCTCGAGG GCCACGACGA CGTGCCCACC TGGTACGTCG ACGTCCCGAC CAACAGCGTC
        CGGGAGCTCC CGGTGCTGCT GCACGGGTGG ACCATGCAGC TGCAGGGCTG GTTGTCGCAG

 481    GTCGTCGCCG TCAAGGCCGG AGCCCAGGAC GTCGCCGCCG GCCTCGTCGA AGGTGCCGAC
        CAGCAGCGGC AGTTCCGGCC TCGGGTCCTG CAGCGGCGGC CGGAGCAGCT TCCACGGCTG
                                                              595
 541    GTCCCGTCCG ACGCCGTGAC CTTCGTCGAG ACCGACGAGA CCCCGCGGAC CATGTTCGAC
        CAGGGCAGGC TGCGGCACTG GAAGCAGCTC TGGCTGCTCT GGGGCGCCTG GTACAAGCTG

 601    GTGATCGGCG GCAACGCCTA CACCATCGGG GGGCGCAGCC GCTGCTCGAT CGGGTTCGCG
        CACTAGCCGC CGTTGCGGAT GTGGTAGCCC CCCGCGTCGG CGACGAGCTA GCCCAAGCGC

 661    GTCAACGGCG GGTTCATCAC CGCCGGCCAC TGCGGCCGCA CCGGCGCCAC CACCGCCAAC
        CAGTTGCCGC CCAAGTAGTG GCGGCCGGTG ACGCCGGCGT GGCCGCGGTG GTGGCGGTTG

 721    CCCACCGGGA CCTTCGCCGG GTCCAGCTTC CCGGGCAACG ACTACGCGTT CGTCCGTACC
        GGGTGGCCCT GGAAGCGGCC CAGGTCGAAG GGCCCGTTGC TGATGCGCAA GCAGGCATGG

 781    GGGGCCGGCG TGAACCTGCT GGCCCAGGTC AACAACTACT CCGGTGGCCG CGTCCAGGTC
        CCCCGGCCGC ACTTGGACGA CCGGGTCCAG TTGTTGATGA GGCCACCGGC GCAGGTCCAG

 841    GCCGGGCACA CCGCGGCCCC CGTCGGCTCG GCCGTGTGCC GGTCCGGGTC GACCACCGGG
        CGGCCCGTGT GGCGCCGGGG GCAGCCGAGC CGGCACACGG CCAGGCCCAG CTGGTGGCCC

0901    TGGCACTGCG GCACCATCAC TGCGCTCAAC TCCTCGGTCA CCTACCCCGA GGGCACCGTC
        ACCGTGACGC CGTGGTAGTG ACGCGAGTTG AGGAGCCAGT GGATGGGGCT CCCGTGGCAG

0961    CGCGGCCTGA TCCGCACCAC CGTCTGCGCC GAGCCCGGCG ACTCCGGTGG CTCGCTGCTC
        GCGCCGGACT AGGCGTGGTG GCAGACGCGG CTCGGGCCGC TGAGGCCACC GAGCGACGAG
```

FIGURE 2B

```
1021    GCCGGCAACC AGGCCCAGGG CGTCACGTCC GGCGGCTCCG GCAACTGCCG CACCGGTGGC
        CGGCCGTTGG TCCGGGTCCC GCAGTGCAGG CCGCCGAGGC CGTTGACGGC GTGGCCACCG

1081    ACCACGTTCT TCCAGCCGGT CAACCCCATC CTCCAGGCGT ACGGCCTGAG GATGATCACC
        TGGTGCAAGA AGGTCGGCCA GTTGGGGTAG GAGGTCCGCA TGCCGGACTC CTACTAGTGG
                              1162
1141    ACGGACTCGG GCAGCAGCCC GGCCCCTGCA CCGACCTCCT GCACCGGCTA CGCCCGCACC
        TGCCTGAGCC CGTCGTCGGG CCGGGGACGT GGCTGGAGGA CGTGGCCGAT GCGGGCGTGG

1201    TTCACCGGGA CCCTCGCGGC CGGCCGGGCC GCCGCCCAGC CCAACGGGTC CTACGTGCAG
        AAGTGGCCCT GGGAGCGCCG GCCGGCCCGG CGGCGGGTCG GGTTGCCCAG GATGCACGTC

1261    GTCAACCGGT CCGGGACCCA CAGCGTGTGC CTCAACGGGC CCTCCGGTGC GGACTTCGAC
        CAGTTGGCCA GGCCCTGGGT GTCGCACACG GAGTTGCCCG GGAGGCCACG CCTGAAGCTG

1321    CTCTACGTGC AGCGCTGGAA CGGCAGCTCC TGGGTGACCG TCGCCCAGAG CACCTCCCCC
        GAGATGCACG TCGCGACCTT GCCGTCGAGG ACCCACTGGC AGCGGGTCTC GTGGAGGGGG

1381    GGCTCCAACG AGACCATCAC CTACCGCGGC AACGCCGGCT ACTACCGCTA CGTGGTCAAC
        CCGAGGTTGC TCTGGTAGTG GATGGCGCCG TTGCGGCCGA TGATGGCGAT GCACCAGTTG
                                                    1486*
1441    GCCGCGTCCG GCTCCGGTGC CTACACCATG GGGCTCACCC TCCCCTGACG TAGCGC
        CGGCGCAGGC CGAGGCCACG GATGTGGTAC CCCGAGTGGG AGGGGACTGC ATCGCG
```

FIGURE 3A

```
   1    AACGAGCCCG CACCACCCGG GAGCGCGAGC GCACCGCCAC GCCTGGCCGA GAAGCTCGAC
        TTGCTCGGGC GTGGTGGGCC CTCGCGCTCG CGTGGCGGTG CGGACCGGCT CTTCGAGCTG

  61    CCCGACCTCC TCGAGGCCAT GGAGCGCGAC CTGGGCCTCG ACGCGGAGGA AGCCGCCGCC
        GGGCTGGAGG AGCTCCGGTA CCTCGCGCTG GACCCGGAGC TGCGCCTCCT TCGGCGGCGG

 121    ACCCTGGCGT TCCAGCACGA CGCAGCCGAG ACCGGCGAGG CCCTCGCCGA AGAGCTCGAC
        TGGGACCGCA AGGTCGTGCT GCGTCGGCTC TGGCCGCTCC GGGAGCGGCT TCTCGAGCTG

 181    GAGGACTTCG CCGGCACCTG GGTCGAGGAC GACGTCCTGT ACGTCGCCAC CACCGACGAG
        CTCCTGAAGC GGCCGTGGAC CCAGCTCCTG CTGCAGGACA TGCAGCGGTG GTGGCTGCTC

 241    GACGCCGTCG AGGAGGTCGA GGGCGAAGGC GCCACGGCCG TCACCGTCGA GCACTCCCTG
        CTGCGGCAGC TCCTCCAGCT CCCGCTTCCG CGGTGCCGGC AGTGGCAGCT CGTGAGGGAC

 301    GCCGACCTCG AGGCCTGGAA GACCGTCCTC GACGCCGCCC TCGAGGGCCA CGACGACGTG
        CGGCTGGAGC TCCGGACCTT CTGGCAGGAG CTGCGGCGGG AGCTCCCGGT GCTGCTGCAC

 361    CCCACCTGGT ACGTCGACGT CCCGACCAAC AGCGTCGTCG TCGCCGTCAA GGCCGGAGCC
        GGGTGGACCA TGCAGCTGCA GGGCTGGTTG TCGCAGCAGC AGCGGCAGTT CCGGCCTCGG

 421    CAGGACGTCG CCGCCGGCCT CGTCGAAGGT GCCGACGTCC CGTCCGACGC CGTGACCTTC
        GTCCTGCAGC GGCGGCCGGA GCAGCTTCCA CGGCTGCAGG GCAGGCTGCG GCACTGGAAG
```

FIG 3B

```
 481    GTCGAGACCG ACGAGACCCC GCGGACCATG TTCGACGTGA TCGGCGGCAA CGCCTACACC
        CAGCTCTGGC TGCTCTGGGG CGCCTGGTAC AAGCTGCACT AGCCGCCGTT GCGGATGTGG

 541    ATCGGGGGGC GCAGCCGCTG CTCGATCGGG TTCGCGGTCA ACGGCGGGTT CATCACCGCC
        TAGCCCCCCG CGTCGGCGAC GAGCTAGCCC AAGCGCCAGT TGCCGCCCAA GTAGTGGCGG

 601    GGCCACTGCG GCCGCACCGG CGCCACCACC GCCAACCCCA CCGGGACCTT CGCCGGGTCC
        CCGGTGACGC CGGCGTGGCC GCGGTGGTGG CGGTTGGGGT GGCCCTGGAA GCGGCCCAGG

 661    AGCTTCCCGG GCAACGACTA CGCGTTCGTC CGTACCGGGG CCGGCGTGAA CCTGCTGGCC
        TCGAAGGGCC CGTTGCTGAT GCGCAAGCAG GCATGGCCCC GGCCGCACTT GGACGACCGG

 721    CAGGTCAACA ACTACTCCGG TGGCCGCGTC CAGGTCGCCG GGCACACCGC GGCCCCCGTC
        GTCCAGTTGT TGATGAGGCC ACCGGCGCAG GTCCAGCGGC CCGTGTGGCG CCGGGGGCAG

 781    GGCTCGGCCG TGTGCCGGTC CGGGTCGACC ACCGGGTGGC ACTGCGGCAC CATCACTGCG
        CCGAGCCGGC ACACGGCCAG GCCCAGCTGG TGGCCCACCG TGACGCCGTG GTAGTGACGC

 841    CTCAACTCCT CGGTCACCTA CCCCGAGGGC ACCGTCCGCG GCCTGATCCG CACCACCGTC
        GAGTTGAGGA GCCAGTGGAT GGGGCTCCCG TGGCAGGCGC CGGACTAGGC GTGGTGGCAG

 901    TGCGCCGAGC CCGGCGACTC CGGTGGCTCG CTGCTCGCCG GCAACCAGGC CCAGGGCGTC
        ACGCGGCTCG GGCCGCTGAG GCCACCGAGC GACGAGCGGC CGTTGGTCCG GGTCCCGCAG

 961    ACGTCCGGCG GCTCCGGCAA CTGCCGCACC GGTGGCACCA CGTTCTTCCA GCCGGTCAAC
        TGCAGGCCGC CGAGGCCGTT GACGGCGTGG CCACCGTGGT GCAAGAAGGT CGGCCAGTTG

1021    CCCATCCTCC AGGCGTACGG CCTGAGGATG ATCACCACGG ACTCGGGCAG CAGCCCGGCC
        GGGTAGGAGG TCCGCATGCC GGACTCCTAC TAGTGGTGCC TGAGCCCGTC GTCGGGCCGG

1081    CCTGCACCGA CCTCCTGCAC CGGCTACGCC CGCACCTTCA CCGGGACCCT CGCGGCCGGC
        GGACGTGGCT GGAGGACGTG GCCGATGCGG GCGTGGAAGT GGCCCTGGGA GCGCCGGCCG

1141    CGGGCCGCCG CCCAGCCCAA CGGGTCCTAC GTGCAGGTCA ACCGGTCCGG GACCCACAGC
        GCCCGGCGGC GGGTCGGGTT GCCCAGGATG CACGTCCAGT TGGCCAGGCC CTGGGTGTCG

1201    GTGTGCCTCA ACGGGCCCTC CGGTGCGGAC TTCGACCTCT ACGTGCAGCG CTGGAACGGC
        CACACGGAGT TGCCCGGGAG GCCACGCCTG AAGCTGGAGA TGCACGTCGC GACCTTGCCG

1261    AGCTCCTGGG TGACCGTCGC CCAGAGCACC TCCCCCGGCT CCAACGAGAC CATCACCTAC
        TCGAGGACCC ACTGGCAGCG GGTCTCGTGG AGGGGGCCGA GGTTGCTCTG GTAGTGGATG

1321    CGCGGCAACG CCGGCTACTA CCGCTACGTG GTCAACGCCG CGTCCGGCTC CGGTGCCTAC
        GCGCCGTTGC GGCCGATGAT GGCGATGCAC CAGTTGCGGC GCAGGCCGAG GCCACGGATG

1381    ACCATGGGGC TCACCCTCCC CTGACGTAGC GC
        TGGTACCCCG AGTGGGAGGG GACTGCATCG CG
```

FIG 4

```
1       TTCGACGTGA TCGGCGGCAA CGCCTACACC ATCGGGGGGC GCAGCCGCTG CTCGATCGGG
        AAGCTGCACT AGCCGCCGTT GCGGATGTGG TAGCCCCCCG CGTCGGCGAC GAGCTAGCCC

61      TTCGCGGTCA ACGGCGGGTT CATCACCGCC GGCCACTGCG GCCGCACCGG CGCCACCACC
        AAGCGCCAGT TGCCGCCCAA GTAGTGGCGG CCGGTGACGC CGGCGTGGCC GCGGTGGTGG

121     GCCAACCCCA CCGGGACCTT CGCCGGGTCC AGCTTCCCGG GCAACGACTA CGCGTTCGTC
        CGGTTGGGGT GGCCCTGGAA GCGGCCCAGG TCGAAGGGCC CGTTGCTGAT GCGCAAGCAG

181     CGTACCGGGG CCGGCGTGAA CCTGCTGGCC CAGGTCAACA ACTACTCCGG TGGCCGCGTC
        GCATGGCCCC GGCCGCACTT GGACGACCGG GTCCAGTTGT TGATGAGGCC ACCGGCGCAG

241     CAGGTCGCCG GGCACACCGC GGCCCCCGTC GGCTCGGCCG TGTGCCGGTC CGGGTCGACC
        GTCCAGCGGC CCGTGTGGCG CCGGGGGCAG CCGAGCCGGC ACACGGCCAG GCCCAGCTGG

301     ACCGGGTGGC ACTGCGGCAC CATCACTGCG CTCAACTCCT CGGTCACCTA CCCCGAGGGC
        TGGCCCACCG TGACGCCGTG GTAGTGACGC GAGTTGAGGA GCCAGTGGAT GGGGCTCCCG

361     ACCGTCCGCG GCCTGATCCG CACCACCGTC TGCGCCGAGC CCGGCGACTC CGGTGGCTCG
        TGGCAGGCGC CGGACTAGGC GTGGTGGCAG ACGCGGCTCG GGCCGCTGAG GCCACCGAGC

421     CTGCTCGCCG GCAACCAGGC CCAGGGCGTC ACGTCCGGCG GCTCCGGCAA CTGCCGCACC
        GACGAGCGGC CGTTGGTCCG GGTCCCGCAG TGCAGGCCGC CGAGGCCGTT GACGGCGTGG

481     GGTGGCACCA CGTTCTTCCA GCCGGTCAAC CCCATCCTCC AGGCGTACGG CCTGAGGATG
        CCACCGTGGT GCAAGAAGGT CGGCCAGTTG GGGTAGGAGG TCCGCATGCC GGACTCCTAC

561     ATCACCACGG ACTCGGGCAG CAGCCCG
        TAGTGGTGCC TGAGCCCGTC GTCGGGC
```

FIG 5

```
1       ATGACACCAC.GCACAGTCAC GCGGGCCCTG GCCGTGGCCA CCGCAGCCGC CACACTCCTG
        TACTGTGGTG CGTGTCAGTG CGCCCGGGAC CGGCACCGGT GGCGTCGGCG GTGTGAGGAC

61      GCAGGCGGCA TGGCCGCCCA GGCC
        CGTCCGCCGT ACCGGCGGGT CCGG
```

FIG 6

```
  1  MTPRTVTRAL AVATAAATLL AGGMAAQA NE PAPPGSASAP PRLAEKLDPD
        1a        1b        1c      2a      2b        2c
 51  LLEAMERDLG LDAEEAAATL AFQHDAAETG EALAEELDED FAGTWVEDDV
        2d        2e        2f        2g              2h
101  LYVATTDEDA VEEVEGEGAT AVTVEHSLAD LEAWKTVLDA ALEGHDDVPT
        2i        2j              2k        2l        2m
151  WYVDVPTNSV VVAVKAGAQD VAAGLVEGAD VPSDAVTFVE TDETPRTM FD
        2n        2o        2p              2q        2r    3a
201  VIGGNAYTIG GRSRCSIGFA VNGGFITAGH CGRTGATTAN PTGTFAGSSF
        3b        3c        3d        3e              3f
251  PGNDYAFVRT GAGVNLLAQV NNYSGGRVQV AGHTAAPVGS AVCRSGSTTG
        3g        3h        3i        3j              3k
301  WHCGTITALN SSVTYPEGTV RGLIRTTVCA EPGDSGGSLL AGNQAQGVTS
        3l        3m        3n              3o        3p
351  GGSGNCRTGG TTFFQPVNPI LQAYGLRMIT TDSGSSP APA PTSCTGYART
        3q        3r        3s        3t      4a      4b
401  FTGTLAAGRA AAQPNGSYVQ VNRSGTHSVC LNGPSGADFD LYVQRWNGSS
        4c        4d        4e              4f        4g
451  WVTVAQSTSP GSNETITYRG NAGYYRYVVN AASGSGAYTM GLTLP
           4h        4i        4j        4k        4l
```

FIG 7

```
1    NEPAPPGSAS APPRLAEKLD PDLLEAMERD.LGLDAEEAAA.TLAFQHDAAE
51   TGEALAEELD EDFAGTWVED DVLYVATTDE DAVEEVEGEG ATAVTVEHSL
101  ADLEAWKTVL DAALEGHDDV PTWYVDVPTN SVVVAVKAGA QDVAAGLVEG
151  ADVPSDAVTF VETDETPRTM FDVIGGNAYT IGGRSRCSIG FAVNGGFITA
201  GHCGRTGATT ANPTGTFAGS SFPGNDYAFV RTGAGVNLLA QVNNYSGGRV
251  QVAGHTAAPV GSAVCRSGST TGWHCGTITA LNSSVTYPEG TVRGLIRTTV
301  CAEPGDSGGS LLAGNQAQGV TSGGSGNCRT GGTTFFQPVN PILQAYGLRM
351  ITTDSGSSPA PAPTSCTGYA RTFTGTLAAG RAAAQPNGSY VQVNRSGTHS
401  VCLNGPSGAD FDLYVQRWNG SSWVTVAQST SPGSNETITY RGNAGYYRYV
451  VNAASGSGAY TMGLTLP
```

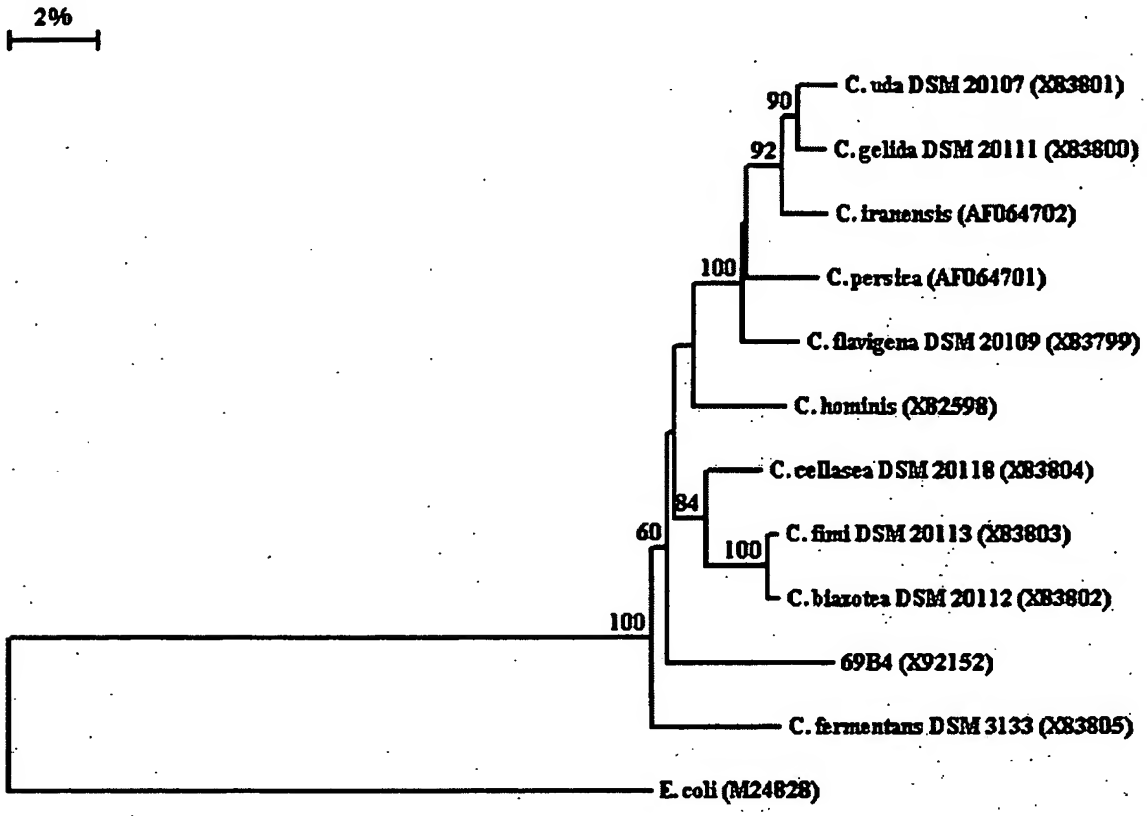
FIG 8

```
  1  FDVIGGNAYT IGGRSRCSIG FAVNGGFITA GHCGRTGATT ANPTGTFAGS
 51  SFPGNDYAFV RTGAGVNLLA QVNNYSGGRV QVAGHTAAPV GSAVCRSGST
101  TGWHCGTITA LNSSVTYPEG TVRGLIRTTV CAEPGDSGGS LLAGNQAQGV
151  TSGGSGNCRT GGTTFFQPVN PILQAYGLRM ITTDSGSSP
```

FIG 9

```
  1  MTPRTVTRAL AVATAAATLL AGGMAAQA
```

FIGURE 10



Phylogenetic tree showing the relationship of strain 69B4 to other Cellulomonas species.

FIG.11

TTGWXCGT_FW: 5' ACNACSGGSTGGCRGTGCGGCAC 3'

FIG 12

GDSGGX_RV: 5'-ANGNGCCGCCGGAGTCNCC-3'

FIG 13

```
          D  G  W   D  C  G   T  I  T  A   L  N  S   S  V  T   Y  P  E  G
   1      ACGACGGCTG GGACTGCGGC ACCATCACTG CGCTCAACTC CTCGGTCACC TACCCCGAGG
          TGCTGCCGAC CCTGACGCCG TGGTAGTGAC GCGAGTTGAG GAGCCAGTGG ATGGGGCTCC
          ·  T  V  R   G  L  I   R  T  T  V   C  A  E   P  G  D   S  G  G  S
  61      GCACCGTCCG CGGCCTGATC CGCACCACCG TCTGCGCCGA GCCCGGCGAC TCCGGTGGCT
          CGTGGCAGGC GCCGGACTAG GCGTGGTGGC AGACGCGGCT CGGGCCGCTG AGGCCACCGA
          ·  L  L  A   G  N  Q   A  Q  G  V   T  S  G   D  S  G   G  S
 121      CGCTGCTCGC CGGCAACCAG GCCCAGGGCG TCACGTCCGG CGACTCCGGC GGCTCAT
          GCGACGAGCG GCCGTTGGTC CGGGTCCCGC AGTGCAGGCC GCTGAGGCCG CCGAGTA
```

FIG 14A

```
                1                                                    50
 AAC23545     (1) --MNHSSR--RTTSLLFTAALAATALVAATTPAS----------------
   PC2053     (1) --MRHTGR-NAIGAAIAASALAFALVPSQAAAN------DTLTERAEAAV
   S34672     (1) --MRLKGRTVAIGSALAASALALSLVPANASSELP----SAETAKADALV
 CAD42808     (1) MVGRHAAR-SRRAALTALGALVLTALPSAASAAPPPVPGPRPAVARTPDA
NP_625129     (1) MVGRHAAR-SRRAALTALGALVLTALPSAASAAPPPVPGPRPAVARTPDA
NP_822175     (1) MVHRHVG--AGCAGLSVLATLVLTGLPAAAAIEPP-GPAPAPSAVQPLGA
 CAD42809     (1) MPHRHRHH-RAVGAAVAATAALLVAGLSGSASAGTAPAGSAPTAAETLRT
NP_628830     (1) MPHRHRHH-RAVGAAVAATAALLVAGLSGSASAGTAPAGSAPTAAETLRT
   P52320     (1) ---MERTT-LRRRALVAGTATVAVGALALAGLTGVASADPAATAAPPVSA
                51                                                   100
 AAC23545    (31) -----AQELALKRDLGLSDAEVAELRAAEAEAVELEEELRDSLGSDFGGV
   PC2053    (42) ADLPAGVLDAMERDLGLSEQEAGLKLVAEHDAALLGETLSADLDAFAGSW
   S34672    (45) EQLPAGMVDAMERDLGVPAAEVGNQLVAEHEAAVLEESLSEDLSGYAGSW
 CAD42808    (50) ATAPARMLSAMERDLRLAPGQAAARPVNEAEAGTRAGMLRNTLGDRFAGA
NP_625129    (50) ATAPARMLSAMERDLRLAPGQAAARLVNEAEAGTRAGMLRNTLGDRFAGA
NP_822175    (48) GNPSTAVLGALQRDLHLTDTQAKTRLVNEMEAGTRAGRLQNALGKHFAGA
 CAD42809    (50) DAAPPALLKAMQRDLGIDRRQAERRLVNEAEAGATAGRLRAALGGDFAGA
NP_628830    (50) DAAPPALLKAMQRDLGLDRRQAERRLVNEAEAGATAGRLRAALGGDFAGA
   P52320    (47) DSLSPGMLAALERDLGLDEDAARSRIANEYRAAAVAAGLEKSLGARYAGA
                101                                                  150
 AAC23545    (76) YLDADT-TEITVAVTDPAAVSRVDADDVTVDVVDFGETALNDFVASLNAI
   PC2053    (92) LAEGT---ELVVATTSEAEAAEITEAGATAEVVDHTLAELDSVKDALDTA
   S34672    (95) IVEGTS--EHVVATTDRAEAAEITAAGATATVVEHSLAELEAVKDILDEA
 CAD42808   (100) WVSGATSAELTVATTDAADTAAIEAQGAKAAVVGRNLAELRAVKEKLDAA
NP_625129   (100) WVSGATSAELTVATTDAADTAAIEAQGAKAAVVGRNLAELRAVKEKLDAA
NP_822175    (98) WVHGAASADLTVATTHATDIPAITAGGATAVVVKTGLDDLKGAKKKLDSA
 CAD42809   (100) WVRGAESGTLTVATTDAGDVAAVEARGAEAKVVRHSLADLDAAKARLDTA
NP_628830   (100) WVRGAESGTLTVATTDAGDVAAIEARGAEAKVVRHSLADLDAAKARLDTA
   P52320    (97) RVSGAK-ATLTVATTDASEAARITEAGARAEVVGHSLDRFEGVKKSLDKA
                151                                                  200
 AAC23545   (125) ADT--ADPKVTGWYTDLESDAVVITTLRGGTPAAEELAERAGLDERAVRI
   PC2053   (139) AES-YDTTDAPVWYVDVTTNGVVLLTSD--VTEAEGFVEAAGVNAAAVDI
   S34672   (143) ATA-NPEDAAPVWYVDVTTNEVVVLASD--VPAAEAFVAASGADASTVRV
 CAD42808   (150) AVR-TRTRQTPVWYVDVKTNRVTVQATG--ASAAAAFVEAAGVPAADVGV
NP_625129   (150) AVR-TRTRQTPVWYVDVKTNRVTVQATG--ASAAAAFVEAAGVPAADVGV
NP_822175   (148) VAHGGTAVNTPVRYVDVRTNRVTLQARS--RAAADALIAAAGVDSGLVDV
 CAD42809   (150) AAG-LNTADAPVWYVDTRTNTVVVEAIR--PAAARSLLTAAGVDGSLAHV
NP_628830   (150) AAG-LNTADAPVWYVDTRTNTVVVEAIR--PAAARSLLTAAGVDGSLAHV
   P52320   (146) ALD-KAPKNVPVWYVDVAANRVVVNAAS--PAAGQAFLKVAGVDRGLVTV
```

FIG 14B

```
                201                                                 250
 AAC23545  (173) VEEDEEPQSLAAIIGGNPYYFGN-YRCSIGFSVRQGSQTGFATAGHCGST
   PC2053  (186) QTSDEQPQAFYDLVGGDAYYMGG-GRCSVGFSVTQGSTPGFATAGHCGTV
   S34672  (190) ERSDESPQPFYDLVGGDAYYIGN-GRCSIGFSVRQGSTPGFVTAGHCGSV
 CAD42808  (197) RVSPDQPRVLEDLVGGDAYYIDDQARCSIGFSVTKDDQEGFATAGHCGDP
NP_625129  (197) RVSPDQPRVLEDLVGGDAYYIDDQARCSIGFSVTKDDQEGFATAGHCGDP
NP_822175  (196) KVSEDRPRALFDIRGGDAYYIDNTARCSVGFSVTKGNQQGFATAGHCGRA
 CAD42809  (197) KNRTERPRTFYDLRGGEAYYINNSSRCSIGFPITKGTQQGFATAGHCDRA
NP_628830  (197) KNRTERPRTFYDLRGGEAYYINNSSRCSIGFPITKGTQQGFATAGHCGRA
   P52320  (193) ARSAEQPRALADIRGGDAYYMNGSGRCSVGFSVTRGTQNGFATAGHCGRV
                251                                                 300
 AAC23545  (222) GTRVS----SPSGTVAGSYFPGRDMGWVRITSADTVTPLVNRYNGGTVTV
   PC2053  (235) GTSTTGYNQAAQGTFEESSFPGDDMAWVSVNSDWNTTPTVNE--GE-VTV
   S34672  (239) GNATTGFNRVSQGTFRGSWFPGRDMAWVAVNSNWTPTSLVRNS-GSGVRV
 CAD42808  (247) GATTTGYNEADQGTFQASTFPGKDMAWVGVNSDWTATPDVKAEGGEKIQL
NP_625129  (247) GATTTGYNEADQGTFQASTFPGKDMAWVGVNSDWTATPDVKAEGGEKIQL
NP_822175  (246) GAPTAGFNEVAQGTVQASVFPGHDMAWVGVNSDWTATPDVAGAAGQNVSI
 CAD42809  (247) GSSTTGANRVAQGTFQGSIFPGRDMAWVATNSSWTATPYVLGAGGQNVQV
NP_628830  (247) GSSTTGANRVAQGTFQGSIFPGRDMAWVATNSSWTATPYVLGAGGQNVQV
   P52320  (243) GTTTNGVNQQAQGTFQGSTFPGRDIAWVATNANWTPRPLVNGYGRGDVTV
                301                                                 350
 AAC23545  (268) TGSQEAATGSSVCRSGATTGWRCGTIQSKNQTVRYAEGTVTGLTRTTACA
   PC2053  (282) SGSTEAAVGASICRSGSTTGWHCGTIQQHNTSVTYPEGTITGVTRTSVCA
   S34672  (288) TGSTQATVGSSICRSGSTTGWRCGTIQQHNTSVTYPQGTITGVTRTSACA
 CAD42808  (297) AGSVEALVGASVCRSGSTTGWHCGTIQQHDTSVTYPEGTVDGLTGTTVCA
NP_625129  (297) AGSVEALVGASVCRSGSTTGWHCGTIQQHDTSVTYPEGTVDGLTETTVCA
NP_822175  (296) AGSVQAIVGAAICRSGSTTGWHCGTVEEHDTSVTYEEGTVDGLTRTTVCA
 CAD42809  (297) TGSTASPVGASVCRSGSTTGWHCGTVTQLNTSVTYQEGTISPVTRTTVCA
NP_628830  (297) TGSTASPVGASVCRSGSTTGWHCGTVTQLNTSVTYQEGTISPVTRTTVCA
   P52320  (293) AGSTASVVGASVCRSGSTTGWHCGTIQQLNTSVTYPEGTISGVTRTSVCA
                351                                                 400

 AAC23545  (318) EGGDSGGPWLTGSQAQGVTSGGTGDCRSGGITFFQPINPLLSYFGLQLVT
   PC2053  (332) EPGDSGGSYISGSQAQGVTSGGSGNCTSGGTTYHQPINPLLSAYGLDLVT
   S34672  (338) QPGDSGGSFISGTQAQGVTSGGSGNCSIGGTTFHQPVNPILSQYGLTLVR
 CAD42808  (347) EPGDSGGPFVSGVQAQGTTSGGSGDCTNGGTTFYQPVNPLLSDFGLTLKT
NP_625129  (347) EPGDSGGPFVSGVQAQGTTSGGSGDCTNGGTTFYQPVNPLLSDFGLTLKT
NP_822175  (346) EPGDSGGSFVSGSQAQGVTSGGSGDCTRGGTTYYQPVNPILSTYGLTLKT
 CAD42809  (347) EPGDSGGSFISGSQAQGVTSGGSGDCRTGGGTFFQPINALLQNYGLTLKT
NP_628830  (347) EPGDSGGSFISGSQAQGVTSGGSGDCRTGGETFFQPINALLQNYGLTLKT
   P52320  (343) EPGDSGGSYISGSQAQGVTSGGSGNCSSGGTTYFQPINPLLQAYGLTLVT
                401                                                 450
 AAC23545  (368) G--------------------------------------------------
   PC2053  (382) G--------------------------------------------------
   S34672  (388) S--------------------------------------------------
 CAD42808  (397) TSAATQTPAPQDNAAA------DAWTAGRVYEVGTTVSYDGVRYRCLQSH
NP_625129  (397) TSAATQTPAPQDNAAA------DAWTAGRVYEVGTTVSYDGVRYRCLQSH
NP_822175  (396) STAPTDTPSDPVDQSG-------VWAAGRVYEVGAQVTYAGVTYQCLQSH
 CAD42809  (397) TGGDDGGGDDGG-----EEPGG-TWAAGTVYQPGDTVTYGGATFRCLQGH
NP_628830  (397) TGGDDGGGDDGGGDDGGEEPGG-TWAAGTVYQPGDTVTYGGATFRCLQGH
   P52320  (393) SGGGTPTDPPTTPPTDSP---GGTWAVGTAYAAGATVTYGGATYRCLQAH
                451          468
 AAC23545  (369) ------------------
   PC2053  (383) ------------------
   S34672  (389) ------------------
 CAD42808  (441) QAQGVGSPASVPALWQRV
NP_625129  (441) QAQGVGSPASVPALWQRV
NP_822175  (439) QAQGVWQPAATPALWQRL
 CAD42809  (441) QAYAGWEPPNVPALWQRV
NP_628830  (446) QAYAGWEPPNVPALWQRV
   P52320  (440) TAQPGWTPADVPALWQRV
```

FIG 15

```
                                 1                                                 50
            69B4 mature     (1) FDVIGGNAYTIGGRSRCSIGFAVN----GGFITAGHCGRTGATT------
Sg-StreptogrisinC mature    (1) ADIRGGDAYYMNGSGRCSVGFSVTRGTQNGFATAGHCGRVGTTTNG--VN
 Sg-StreptogrisinBmature    (1) --ISGGDAIYSST-GRCSLGFNVRSGSTYYFLTAGHCTDGATTWWANSAR
 Sg-StreptogrisinAmature    (1) --IAGGEAITTGG-SRCSLGFNVSVNGVAHALTAGHCTNISASWS-----
 Sg-StreptogrisinDmature    (1) --IAGGDAIWGSG-SRCSLGFNVVKGGEPYFLTAGHCTESVTSWSD-TQG
              Consensus     (1)   IAGGDAIY  G SRCSLGFNV  G   YFLTAGHCT  GTTW

                                51                                                100
             Asp mature    (41) ANPTGTFAGSSFPGNDYAFVRTGAGVNLLAQVNNYSGGRVQVAGHTAAPV
Sg-StreptogrisinC mature   (49) QQAQGTFQGSTFPGRDIAWVATNANWTPRPLVNGYGRGDVTVAGSTASVV
 Sg-StreptogrisinBmature   (48) TTVLGTTSGSSFPNNDYGIVRYTNTTIPKDGTVGG----QDITSAANATV
 Sg-StreptogrisinAmature   (43) ---IGTRTGTSFPNNDYGIIRHSNPAAADGRVYLYNGSYQDITTAGNAFV
 Sg-StreptogrisinDmature   (47) GSEIGANEGSSFPENDYGLVKYTSDTAHPSEVNLYDGSTQAITQAGDATV
              Consensus    (51)    IGT  GSSFP NDYGIVRYTA       VN Y G  Q IT AG A V
                               101                                               150
             Asp mature    (91) GSAVCRSGSTTGWHCGTITALNSSVTYPEG-TVRGLIRTTVCAEPGDSGG
Sg-StreptogrisinC mature   (99) GASVCRSGSTTGWHCGTIQQLNTSVTYPEG-TISGVTRTSVCAEPGDSGG
 Sg-StreptogrisinBmature   (94) GMAVTRRGSTTGTHSGSVTALNATVNYGGGDVVYGMIRTNVCAEPGDSGG
 Sg-StreptogrisinAmature   (90) GQAVQRSGSTTGLRSGSVTGLNATVNYGSSGIVYGMIQTNVCAEPGDSGG
 Sg-StreptogrisinDmature   (97) GQAVTRSGSTTQVHDGEVTALDATVNYGNGDIVNGLIQTTVCAEPGDSGG
              Consensus   (101) G AV RSGSTTG H GSVTALNATVNYG G IV GLIRTTVCAEPGDSGG
                               151                                               200
             Asp mature   (140) SLLAGNQAQGVTSGGSGNCRTGGTTFFQPVNPILQAYGLRMITTDSGSSP
Sg-StreptogrisinC mature  (148) SYISGSQAQGVTSGGSGNCSSGGTTYFQPINPLLQAYGLTLVTSGGGTPT
 Sg-StreptogrisinBmature  (144) PLYSGTRAIGLTSGGSGNCSSGGTTFFQPVTEALSAYGVSVY--------
 Sg-StreptogrisinAmature  (140) SLFAGSTALGLTSGGSGNCRTGGTTFYQPVTEALSAYGATVL--------
 Sg-StreptogrisinDmature  (147) ALFAGDTALGLTSGGSGDCSSGGTTFFQPVPEALAAYGAEIG--------
              Consensus   (151) SLFAGS ALGLTSGGSGNCSSGGTTFFQPV EALSAYGLTVI
                               201                                               250
             Asp mature   (190) --------------------------------------------------
Sg-StreptogrisinC mature  (198) DPPTTPPTDSPGGTWAVGTAYAAGATVTYGGATYRCLQAHTAQPGWTPAD
 Sg-StreptogrisinBmature  (186) --------------------------------------------------
 Sg-StreptogrisinAmature  (182) --------------------------------------------------
 Sg-StreptogrisinDmature  (189) --------------------------------------------------
              Consensus   (201)
                               251
             Asp mature   (190) --------
Sg-StreptogrisinC mature  (248) VPALWQRV
 Sg-StreptogrisinBmature  (186) --------
 Sg-StreptogrisinAmature  (182) --------
 Sg-StreptogrisinDmature  (189) --------
              Consensus   (251)
```

FIG 16

```
                    1                                                  50
Aqualysin I     (1) ----MRKTYWLMALFAVLVLGGCQMASRSDPTPTLAEAFWPKEAPVYGLD
      69B4      (1) MTPRTVTRALAVATAAATLLAGGMAAQANEPAPPGSASAPPRLAEKLDPD
  Consensus     (1)           MA  A  LLAG   A   DP P  A A  PK A     D
                   51                                                 100
Aqualysin I    (47) DPEAIPGRYIVVFKKGKGQSLLQGGITTLQARLAPQGVVVTQAYTGALQG
      69B4     (51) LLEAMERDLGLDAEEAAATLAFQHDAAETGEALAEE---LDEDFAGTWVE
  Consensus    (51)   EAI     L    A A    Q         LA     L   F G
                  101                                                 150
Aqualysin I    (97) FAAEMAPQALEAFRQSPDVEFIEADKVVRAWATQSPAPWGLDRIDQRDLP
      69B4     (98) DDVLYVATTDEDAVEEVEGEGATAVTVEHSLADLEAWKTVLDAALEGHDD
  Consensus   (101)          E      D E   A  V  A A        LD
                  151                                                 200
Aqualysin I   (147) LSNSYTYTATGRGVNVYYVIDTGIRTTHREFGGRARVGYDALGGNGQDCNG
      69B4    (148) VPTWYVDVPTNS--VVVAVKAGAQDVAAGLVEGADVPSDAVT--FVETDE
  Consensus   (151) L   Y    T     V  I  G           A V  DAL     D
                  201                                                 250
Aqualysin I   (197) HGTHVAGTIGGVTYGVAKAVNLYAVRVLDCNGSGSTSGVIAGVDWVTRNH
      69B4    (194) TPRTMFDVIGGNAYTIGGRS--------RCSIGFAVNGGFITAGHCGRTG
  Consensus   (201)    M   IGG  Y IA            C    A  G        R
                  251                                                 300
Aqualysin I   (247) RRPAVANMSLGGGVSTALDNAVKNSIAAGVVYAVAAGNDNANACNYSPAR
      69B4    (236) ATTANPTGTFAGSSFPGNDYAFVRTGAG--------VNLLAQVNNYSGGR
  Consensus   (251)   A    S AG    A D A   S AA         N  AN  NYS AR
                  301                                                 350
Aqualysin I   (297) VAEALTVGATTSSDARASFSNYGSCVDLFAPGASIPSAWYTSDTATQTLN
      69B4    (278) VQVAGHTAAPVGSAVCRSGSTTGWHCGTIT--ALNSSVTYPEGTVRGLIR
  Consensus   (301) V  A   AA   S    S S  G  .      A   S  Y  T    I
                  351                                                 400
Aqualysin I   (347) GTSMATPHVAGVAALYLEQNPSATPASVASAILNGATTGRLSGIGSGSPN
      69B4    (326) TTVCAEPGDSGGSLLAGNQAQGVTSGGSGNCRTGGTTFFQPVNPILQAYG
  Consensus   (351)  T  A P  AG A L   Q    T A  A     G T          A
                  401                                                 450
Aqualysin I   (397) RLLYSLLSSGSGSTAPCTSCSYYTGSLSG---PGDYNFQPNGTYYYSP-A
      69B4    (376) LRMITTDS-GSSPAPAPTSCTGYARTFTGTLAAGRAAAQPNGSYVQVNRS
  Consensus   (401)   L S  S GS      TSCS Y  S SG    G    QPNGSY     A
                  451                                                 500
Aqualysin I   (443) GTHRAWLRGPAGTDFDLYLWRWDGSRWLTVGSSTGPTSEESLSYSGTAGY
      69B4    (425) GTHSVCLNGPSGADFDLYVQRWNGSSWVTVAQSTSPGSNETITYRGNAGY
  Consensus   (451) GTH   L GPAG DFDLYL RW GS WLTVA ST P S ESISY G AGY
                  501                521
Aqualysin I   (493) YLWRIYAYSGSGMYEFWLQRP
      69B4    (475) YRYVVNAASGSGAYTMGLTLP
  Consensus   (501) Y W I A SGSG Y   L  P
```

FIGURE 17

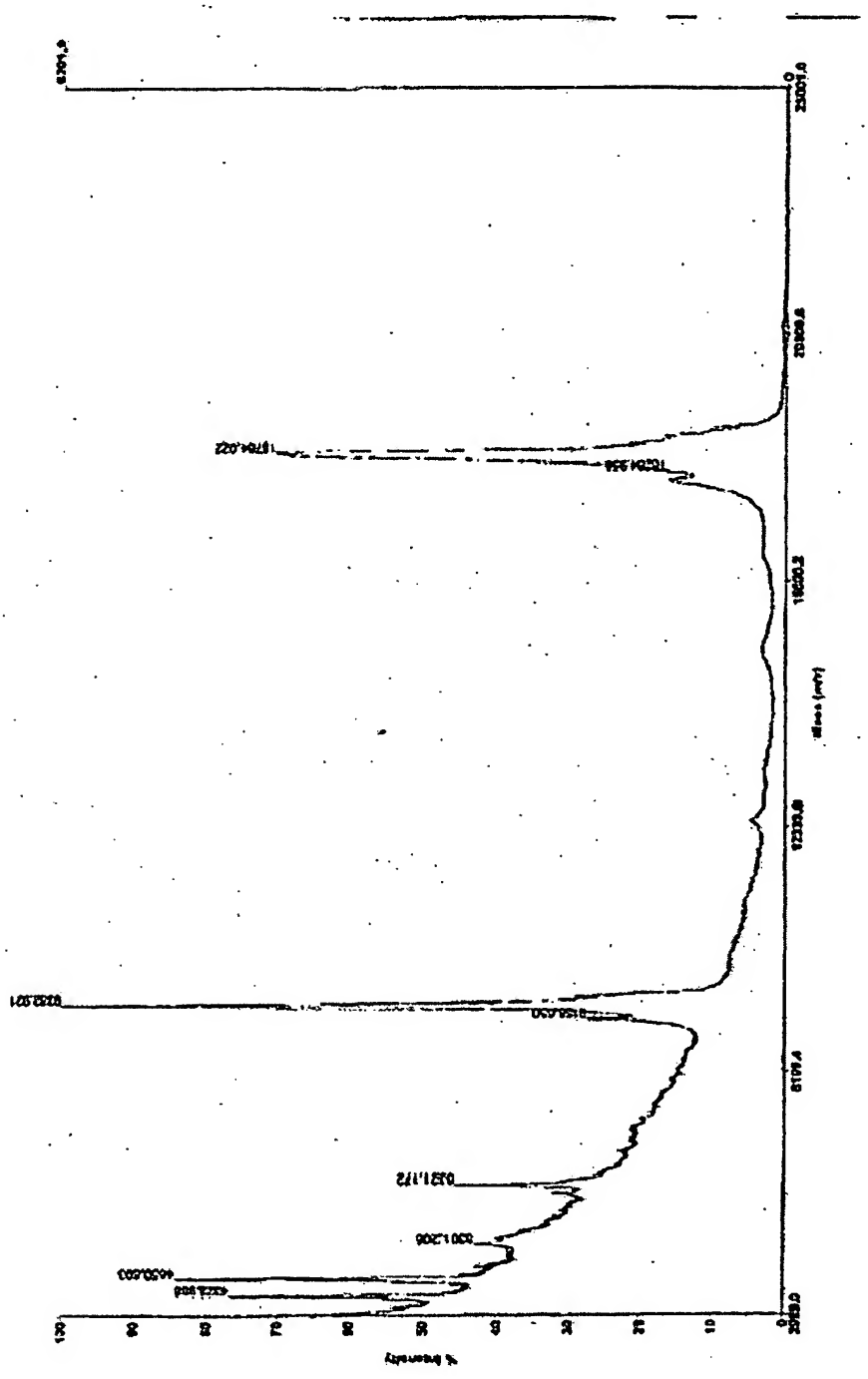
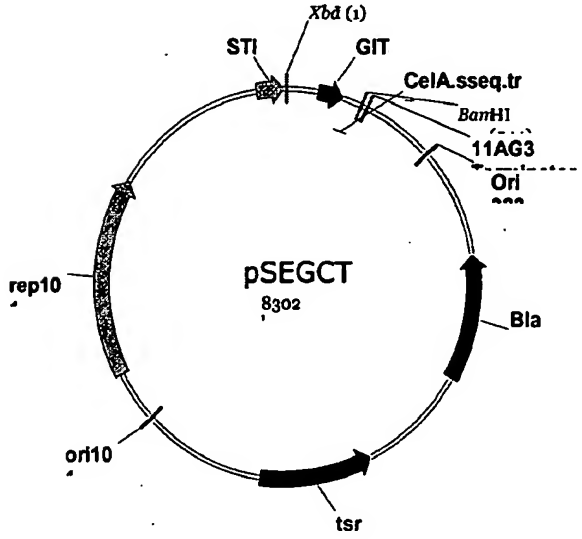
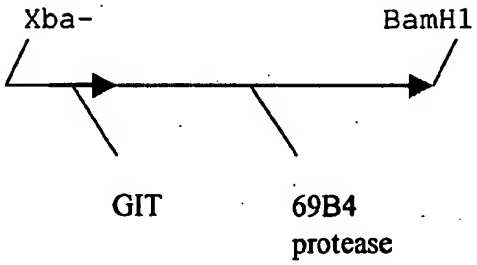


FIGURE 18


Xbd (1)
STI
GIT
CelA.sseq.tr
BamH1
11AG3
Ori
pSEGCT
8302
rep10
Bla
ori10
tsr


Xba-
BamH1
GIT
69B4
protease

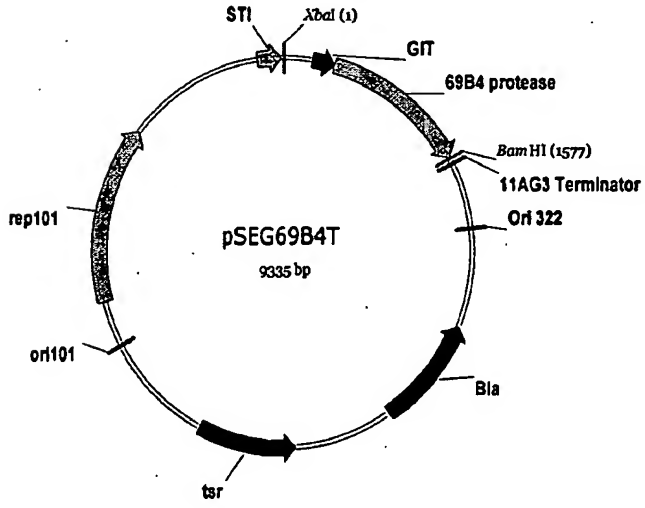

STI
XbaI (1)
GIT
69B4 protease
BamH1 (1577)
11AG3 Terminator
Ori 322
rep101
pSEG69B4T
9335 bp
ori101
Bla
tsr

FIG 19

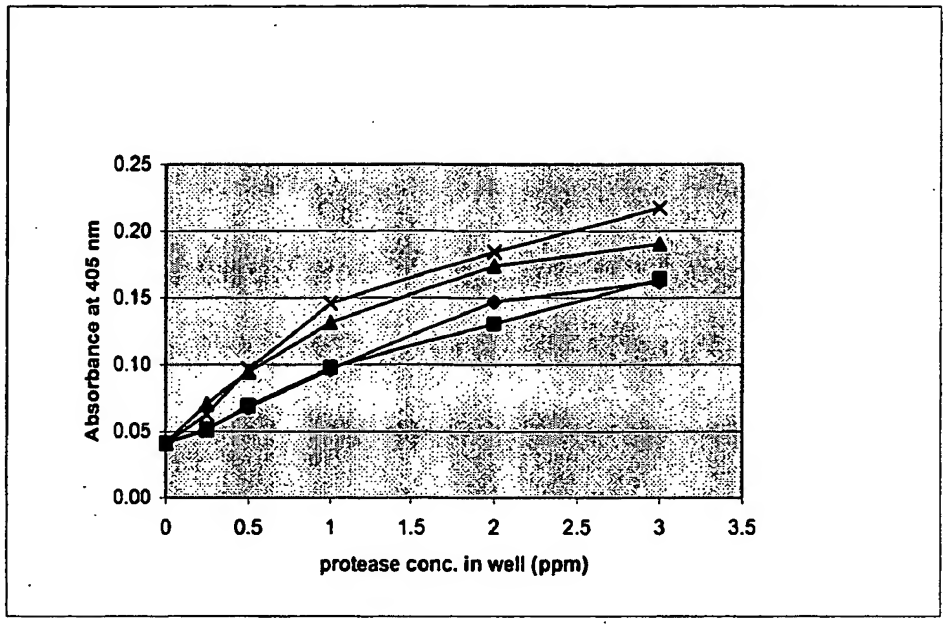

0.25
0.20
0.15
0.10
0.05
0.00
Absorbance at 405 nm
0
0.5
1
1.5
2
2.5
3
3.5
protease conc. in well (ppm)

FIG 20

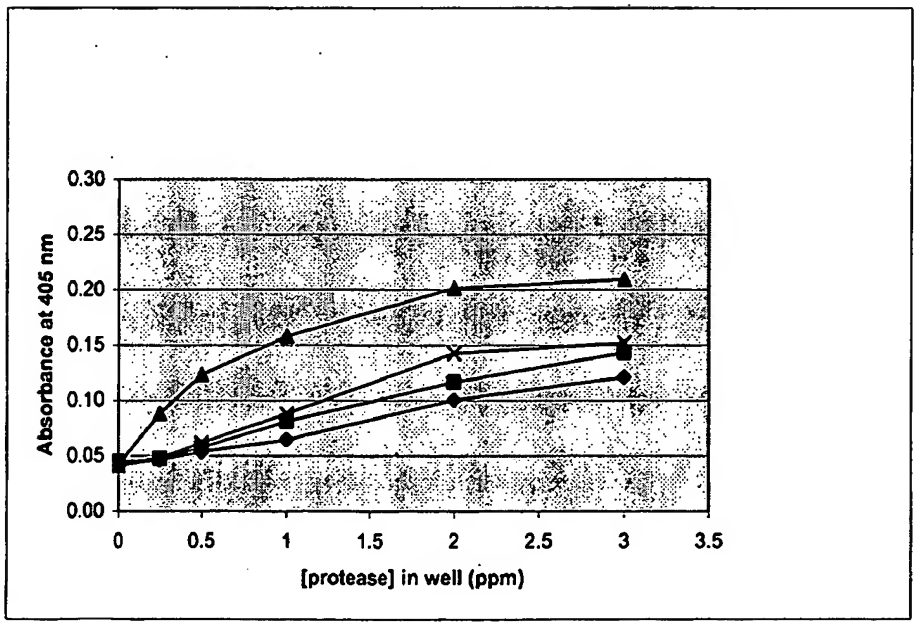

0.30
0.25
0.20
0.15
0.10
0.05
0.00
Absorbance at 405 nm
0
0.5
1
1.5
2
2.5
3
3.5
[protease] in well (ppm)

FIG 21

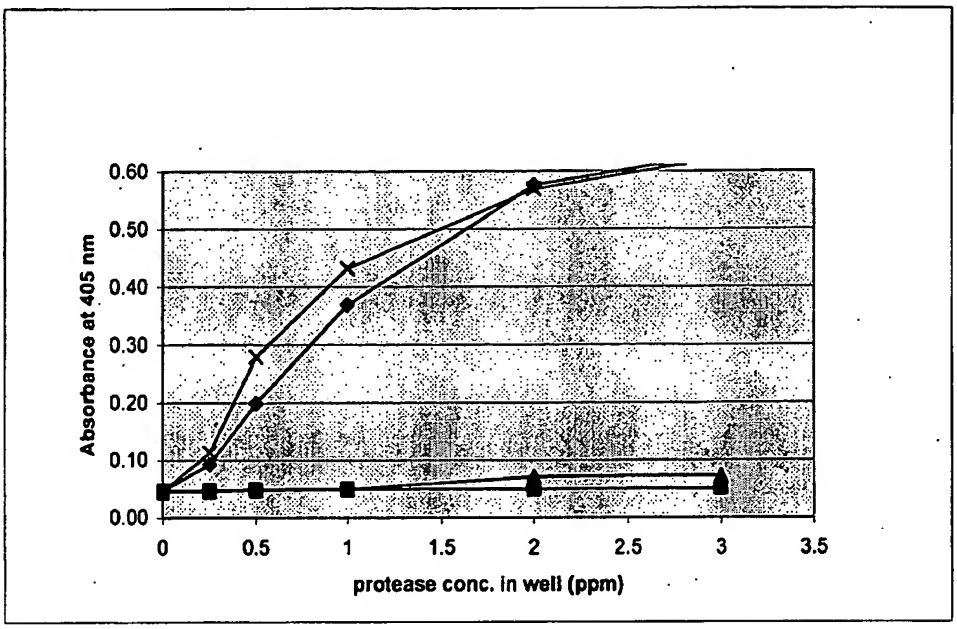

0.60
0.50
0.40
0.30
0.20
0.10
0.00
Absorbance at 405 nm
0
0.5
1
1.5
2
2.5
3
3.5
protease conc. in well (ppm)

FIG 22

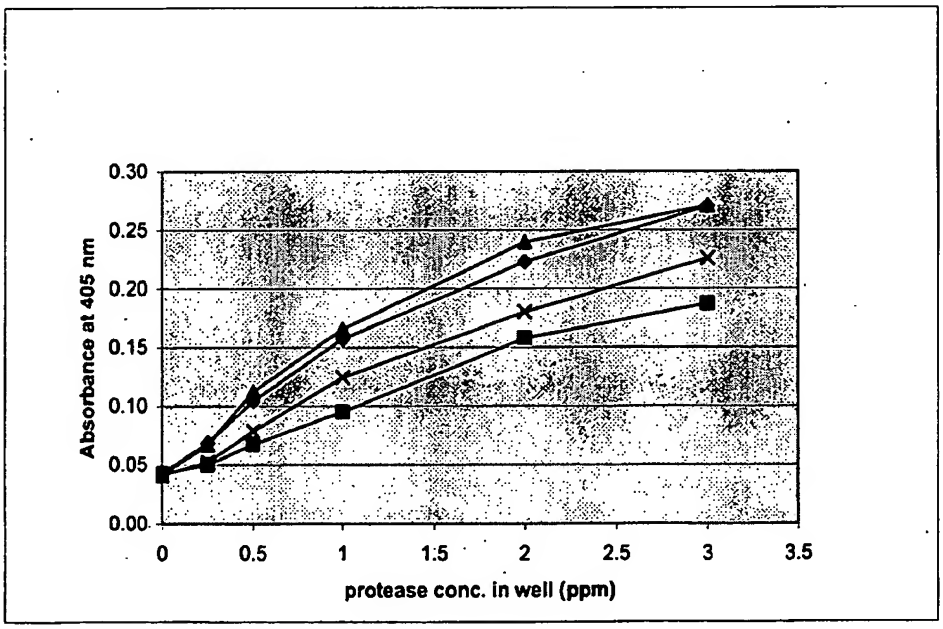

0.30
0.25
0.20
0.15
0.10
0.05
0.00
Absorbance at 405 nm
0
0.5
1
1.5
2
2.5
3
3.5
protease conc. in well (ppm)

FIG 23

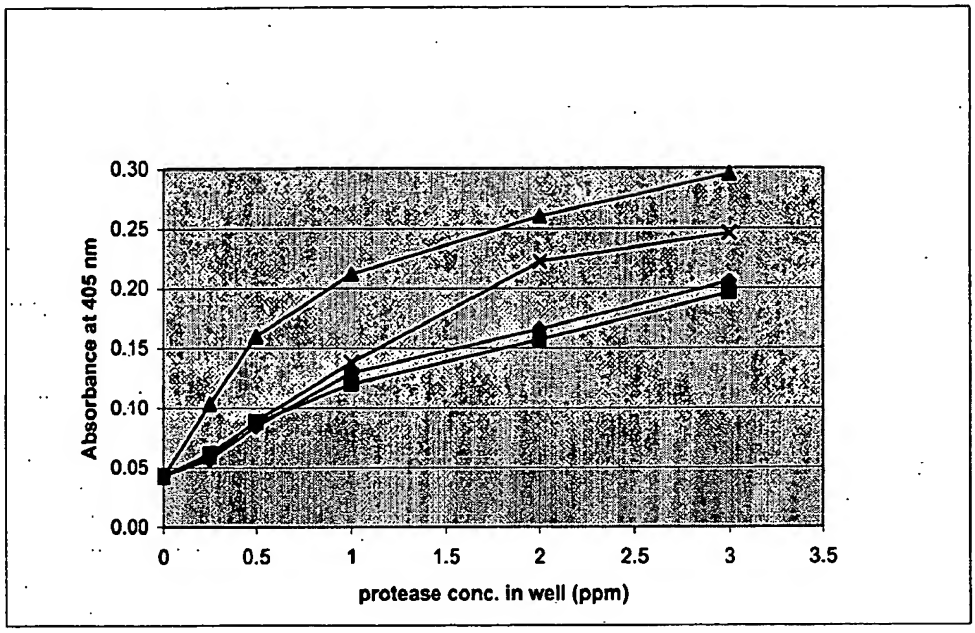

Absorbance at 405 nm
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
0.5
1
1.5
2
2.5
3
3.5
protease conc. in well (ppm)

FIG 24

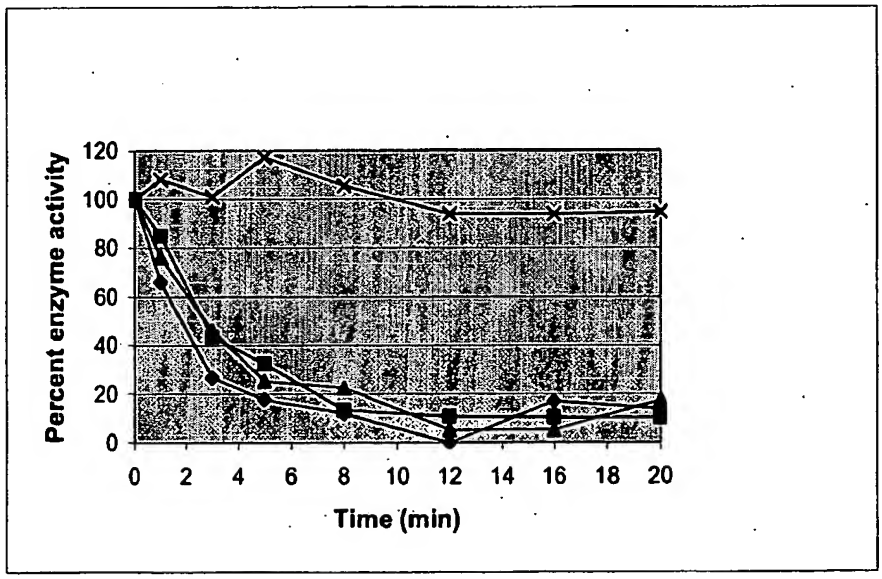

Percent enzyme activity
120
100
80
60
40
20
0
0
2
4
6
8
10
12
14
16
18
20
Time (min)

FIG 25

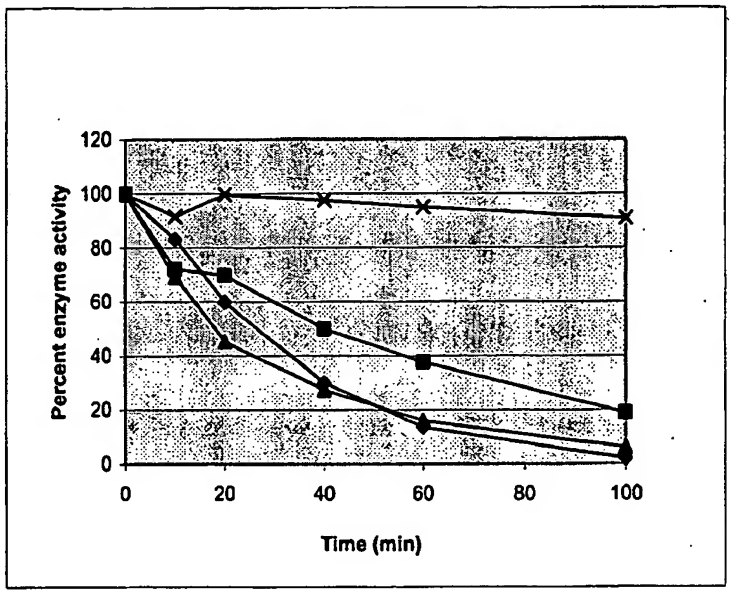

Percent enzyme activity
120
100
80
60
40
20
0
0
20
40
60
80
100
Time (min)

FIG 26

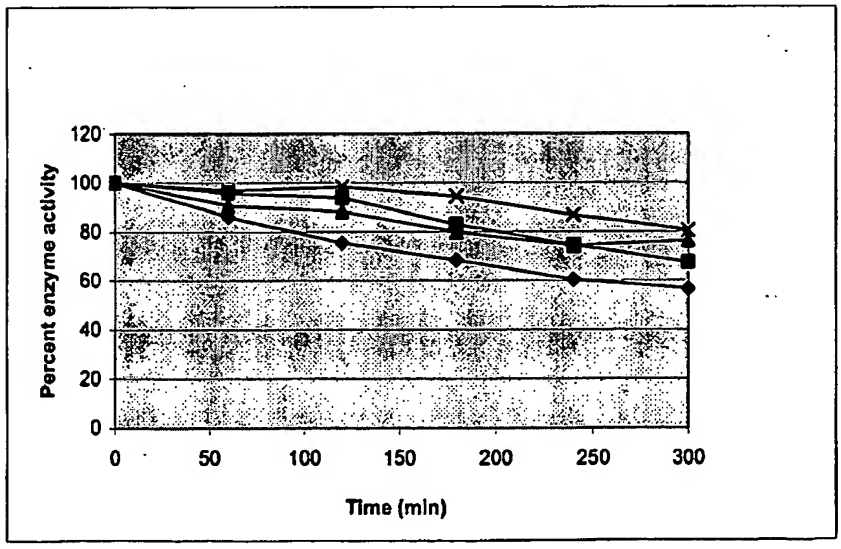

Percent enzyme activity
120
100
80
60
40
20
0
0
50
100
150
200
250
300
Time (min)

FIG 27

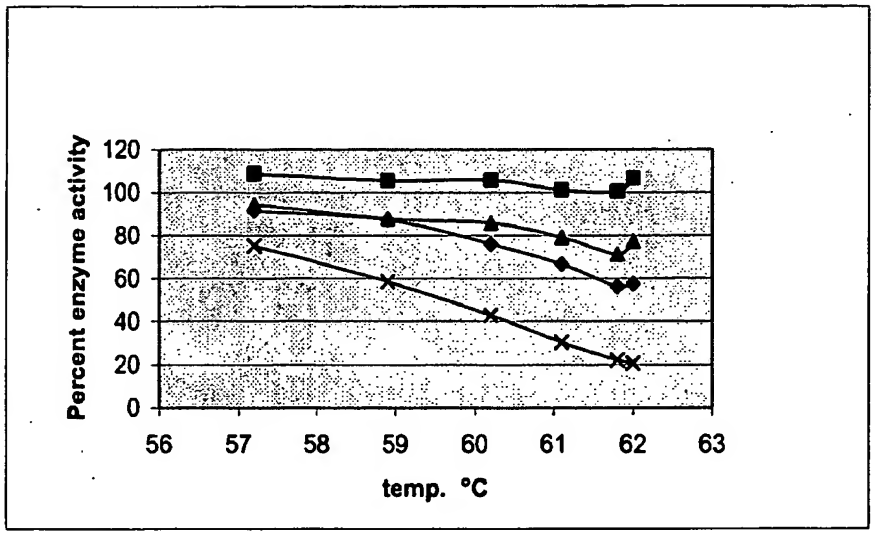

Percent enzyme activity
120
100
80
60
40
20
0
56
57
58
59
60
61
62
63
temp. °C

FIG 28

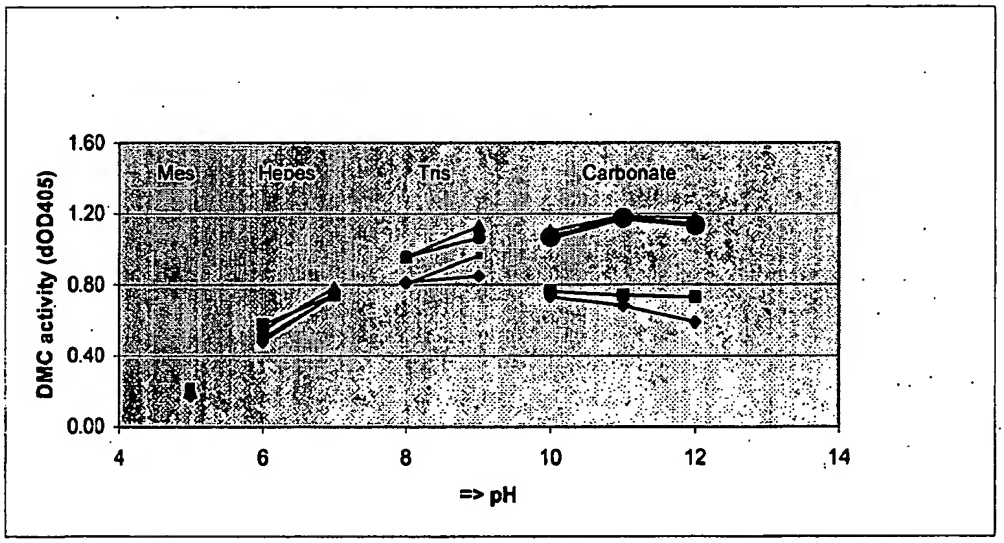

DMC activity (dOD405)
1.60
1.20
0.80
0.40
0.00
Mes
Hepes
Tris
Carbonate
4
6
8
10
12
14
=> pH

FIG 29

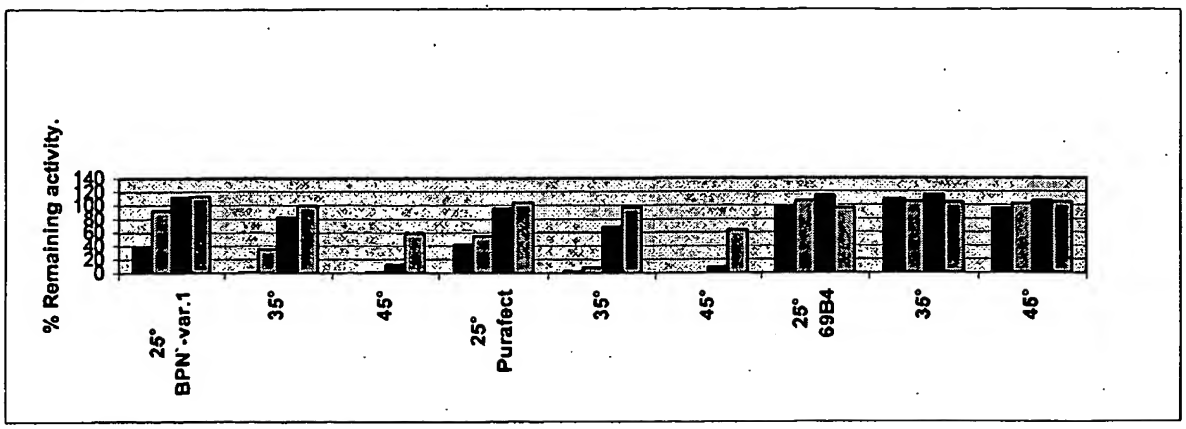

% Remaining activity.
25° BPN'-var.1
35°
45°
25° Purafect
35°
45°
25° 69B4
35°
45°

FIG 30

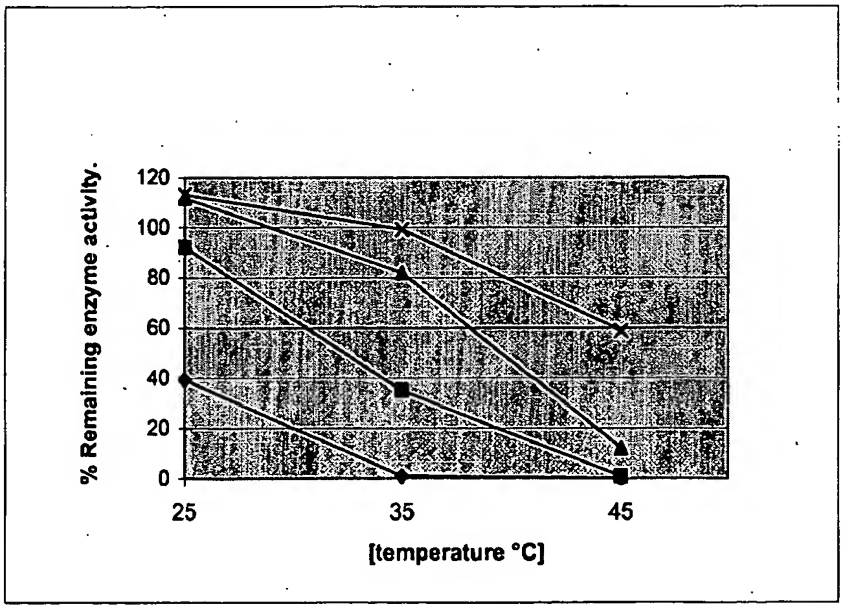

% Remaining enzyme activity.
120
100
80
60
40
20
0
25
35
45
[temperature °C]

FIG 31

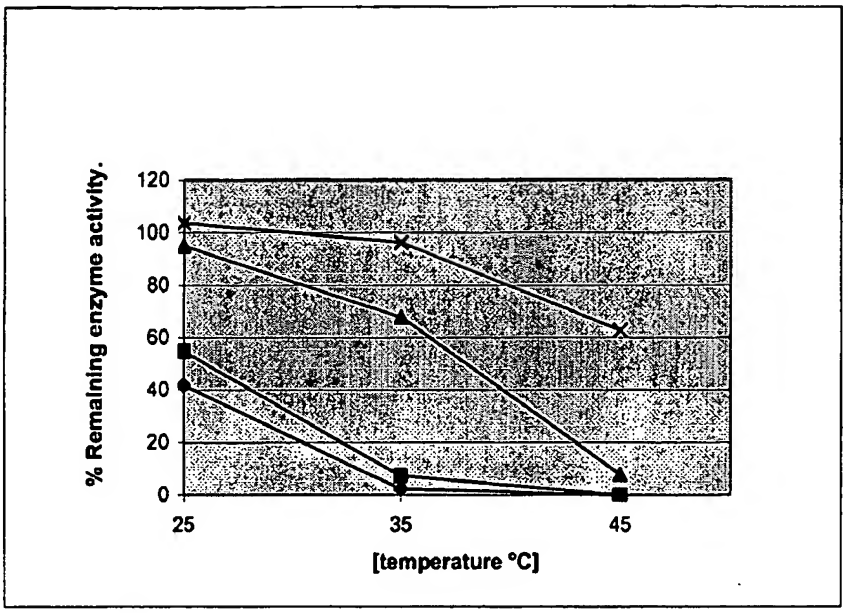
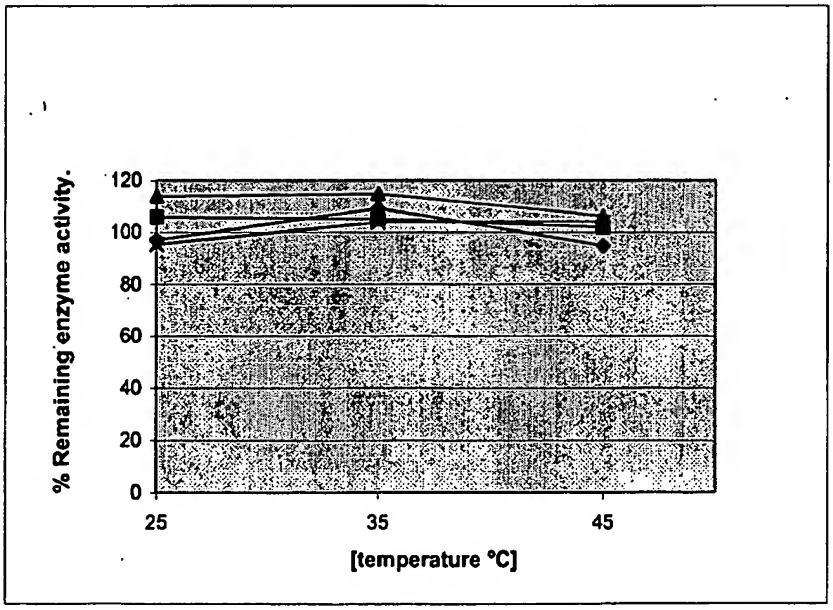

% Remaining enzyme activity.
120
100
80
60
40
20
0
25
35
45
[temperature °C]

FIG 32


% Remaining enzyme activity.
120
100
80
60
40
20
0
25
35
45
[temperature °C]

Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US04/039066

International filing date: 19 November 2004 (19.11.2004)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/523,609
Filing date: 19 November 2003 (19.11.2003)

Date of receipt at the International Bureau: 17 January 2005 (17.01.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse